An Elementary Introduction to the Wiener Process and Stochastic Integrals

Tamás Szabados
Technical University of Budapest

August 23, 1994

Dedicated to Pál Révész on the occasion of his 60th birthday

Abstract

An elementary construction of the Wiener process is discussed, based on a proper sequence of simple symmetric random walks that uniformly converge on bounded intervals, with probability 1. This method is a simplification of F.B. Knight's and P. Révész's. The same sequence is applied to give elementary (Lebesgue-type) definitions of Itô and Stratonovich sense stochastic integrals and to prove the basic Itô formula. The resulting approximating sums converge with probability 1. As a by-product, new elementary proofs are given for some properties of the Wiener process, like the almost sure non-differentiability of the sample-functions. The purpose of using elementary methods almost exclusively is twofold: first, to provide an introduction to these topics for a wide audience; second, to create an approach well-suited for generalization and for attacking otherwise hard problems.

Contents

1991 *Mathematics Subject Classification.* Primary 60J65, 60H05. Secondary 60J15, 60F15.

Key words and phrases. Brownian motion, Wiener process, random walks, stochastic integrals, Itô formula.

1 Introduction

The *Wiener process* is undoubtedly one of the most important stochastic processes, both in the theory and in the applications. Originally it was introduced as a mathematical model of *Brownian motion*, a random zigzag motion of microscopic particles suspended in liquid, discovered by the English botanist Brown in 1827. An amazing number of first class scientists like Bachelier, Einstein, Smoluchowski, Wiener, and Lévy, to mention just a few, contributed to the theory of Brownian motion. In the course of the evolution of probability theory it became clear that the Wiener process is a basic tool for many limit theorems and also a natural model of many phenomena involving randomness, like noise, random fluctuations or perturbations.

The Wiener process is a natural model of Brownian motion. It describes a random, but continuous motion of a particle, subjected to the influence of a large number of chaotically moving molecules of the liquid. Any displacement of the particle over an interval of time as a sum of many almost independent small influences is normally distributed with expectation zero and variance proportional to the length of the time interval. Displacements over disjoint time intervals are independent.

The most basic types of *stochastic integrals* were introduced by K. Itô and R. L. Stratonovich as tools for investigating stochastic differential equations, that is, differential equations containing random functions. Not surprisingly, the Wiener process is one of the corner stones the theory of stochastic integrals and differential equations was built on.

Stochastic differential equations are applied under similar conditions as differential equations in general. The advantage of the stochastic model is that it can accommodate noise or other randomly changing input and effects, which is a necessity in many applications. When solving a stochastic differential equation one has to integrate a function with respect to the increments of a stochastic process like the Wiener process. In such a case the classical methods of integration cannot be applied directly because of the "strange" behaviour of the increments of the Wiener and similar processes.

A main purpose of this paper is to provide an elementary introduction to the aforementioned topics. The discussion of the Wiener process is based on a nice, natural construction of P. Révész [6, Section 6.2], which is essentially a simplified version of F.B. Knight's [4, Section 1.3]. We use a proper sequence of simple random walks that converge to the Wiener

process. Then an elementary definition and discussion of stochastic integrals is given, based on [8], which uses the same sequence of random walks.

The level of the paper is (hopefully) available to any good student who has taken a usual calculus sequence and an introductory course in probability. Our general reference will be W. Feller's excellent, elementary textbook [2]. Anything that goes beyond the material of that book will be discussed here in detail. I would like to convince the reader that these important and widely used topics are natural and feasible supplements to a strong introductory course in probability; this way a much wider audience could get acquainted with them. However, I have to warn the non-expert reader that "elementary" is not a synonym of "easy" or "short".

To encourage the reader it seems worthwhile to emphasize a very useful feature of elementary approaches: in many cases, elementary methods are easier to generalize or to attack otherwise hard problems.

2 Random Walks

The simplest (and crudest) model of Brownian motion is a *simple symmetric random walk* in one dimension, hereafter *random walk* for brevity.

A particle starts from the origin and steps one unit either to the left or to the right with equal probabilities 1/2, in each unit of time. Mathematically, we have a sequence $X_1, X_2, \ldots$ of independent and identically distributed random variables with

$$\mathbf{P}\left\{X_n = 1\right\} = \mathbf{P}\left\{X_n = -1\right\} = 1/2 \ (n = 1, 2, \ldots),$$

and the position of the particle at time n (that is, the random walk) is given by the partial sums

$$S_0 = 0, \ S_n = X_1 + X_2 + \cdots + X_n \ (n = 1, 2, \ldots). \tag{1}$$

The notation $X(n)$ and $S(n)$ will be used instead of X_n and S_n where it seems to be advantageous.

A bit of terminology: a *stochastic process* is a collection $Z(t)$ $(t \in T)$ of random variables defined on a *sample space* Ω. Usually T is a subset of the real line and t is called "time". An important concept is that of a *sample-function*, that is, a randomly selected path of a stochastic process. A sample-function of a stochastic process $Z(t)$ can be denoted by $Z(t; \omega)$, where $\omega \in \Omega$ is fixed, but the "time" t is not.

To visualize the graph of a sample-function of the random walk one can use a broken line connecting the vertices (n, S_n), $n = 1, 2, \ldots$ (Figure 1). This way the sample-functions are extended from the set of the non-negative integers to continuous functions on the interval $[0, \infty)$:

$$S(t) = S_n + (t - n) X_{n+1} \ (n \le t < n + 1; \ n = 0, 1, 2, \ldots). \tag{2}$$

It is easy to evaluate the expectation and variance of S_n:

$$\mathbf{E}(S_n) = \sum_{k=1}^{n} \mathbf{E}(X_k) = 0, \quad \mathbf{Var}(S_n) = \sum_{k=1}^{n} \mathbf{E}(X_k^2) = n. \tag{3}$$

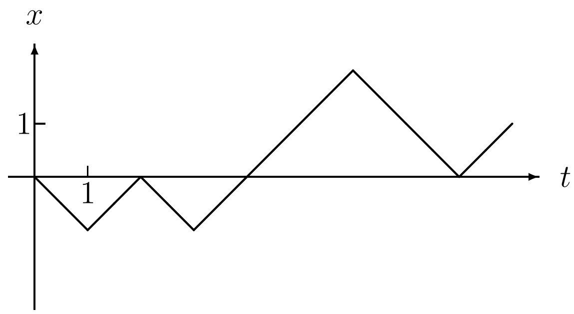

Figure 1: The graph of a sample-function of $S(t)$.

The distribution of S_n is a linearly transformed symmetric binomial distribution [2, Section III,2]. Each path (broken line) of length n has probability $1/2^n$. The number of paths going to the point (n, r) from the origin is equal to the number of choosing $(n + r)/2$ steps to the right out of n steps. Consequently,

$$\mathbf{P}\{S_n = r\} = \binom{n}{(n+r)/2}\frac{1}{2^n} \quad (|r| \le n).$$

The binomial coefficient here is considered to be zero when $n + r$ is not divisible by 2. Equivalently, $S_n = 2B_n - n$, where B_n is a symmetric $(p = 1/2)$ binomial random variable, $\mathbf{P}\{B_n = k\} = \binom{n}{k}2^{-n}$.

An elementary computation shows that for n large, the binomial distribution can be approximated by the normal distribution, see [2, Section VII,2]. What is shown there that for even numbers $n = 2\nu$ and $r = 2k$, if $n \to \infty$ and $|r| < K_n = o(n^{2/3})$, one has

$$\mathbf{P}\{S_n = r\} = \binom{n}{(n+r)/2}\frac{1}{2^n} = \binom{2\nu}{\nu+k}\frac{1}{2^{2\nu}} \sim \frac{1}{\sqrt{\pi\nu}}e^{-k^2/\nu} = 2h\ \phi(rh), \tag{4}$$

where $h = 1/\sqrt{n}$ and $\phi(x) = (1/\sqrt{2\pi})e^{-x^2/2}$ $(-\infty < x < \infty)$, the standard normal density function. Note that for odd numbers $n = 2\nu + 1$ and $r = 2k + 1$ (4) can be proved similarly as for even numbers.

Here and later we adopt the usual notations $a_n \sim b_n$ for $\lim_{n\to\infty} a_n/b_n = 1$ (a_n and b_n are *asymptotically equal*), and $a_n = o(b_n)$ for $\lim_{n\to\infty} a_n/b_n = 0$.

Equation (4) easily implies a special case of the *central limit theorem* and of the *large deviation theorem*, [2, Sections VII,3 and 6]):

Theorem 1 *(a) For any real x fixed and $n \to \infty$ we have*

$$\mathbf{P}\left\{S_n/\sqrt{n} \le x\right\} \to \Phi(x),$$

where $\Phi(x) = (1/\sqrt{2\pi})\int_{-\infty}^{x} e^{-u^2/2}\,du$ $(-\infty < x < \infty)$ is the standard normal distribution function.

(b) If $n \to \infty$ and $x_n \to \infty$ so that $x_n = o(n^{1/6})$, then

$$\mathbf{P}\left\{S_n/\sqrt{n} \ \geq x_n\right\} \sim 1 - \Phi(x_n),$$

$$\mathbf{P}\left\{S_n/\sqrt{n} \ \leq -x_n\right\} \sim \Phi(-x_n) = 1 - \Phi(x_n).\square$$

For us the most essential statement of the theorem is that when x_n goes to infinity (slower than $n^{1/6}$), then the two sides of (6) tend to zero equally fast, in fact very fast. For, to estimate $1 - \Phi(x)$ for x large, one can use the following inequality, see [2, Section VII,1],

$$1 - \Phi(x) < \frac{1}{x\sqrt{2\pi}} e^{-x^2/2} \quad (x > 0). \tag{5}$$

Thus fixing an $\epsilon > 0$, say $\epsilon = 1/2$, there exists an integer $n_0 > 0$ such that

$$\mathbf{P}\left\{\left|\frac{S_n}{\sqrt{n}}\right| \geq x_n\right\} \leq \frac{2(1+\epsilon)}{x_n\sqrt{2\pi}} e^{-x_n^2/2} \leq e^{-x_n^2/2}, \tag{6}$$

for $n \geq n_0$, whenever $x_n \to \infty$ and $x_n = o(n^{1/6})$ as $n \to \infty$. It is important to observe that though S_n can take on every integer from $-n$ to n with positive probability, the event $\{|S_n| > x_n\sqrt{n}\}$ is negligible as $n \to \infty$.

But what can we do if n does not go to ∞, or if the condition $x_n = o(n^{1/6})$ does not hold? Then a simple, but still powerful tool, *Chebyshev's inequality* can be used. A standard form of Chebyshev's inequality [2, Section IX,6] is

$$\mathbf{P}\left\{|X - \mathbf{E}(X)| \geq t\right\} \leq \frac{\mathbf{Var}(X)}{t^2},$$

for any $t > 0$, supposing $\mathbf{Var}(X)$ is finite. An other form that can be proved similarly is

$$\mathbf{P}\left\{|X| \geq t\right\} \leq \frac{\mathbf{E}(|X|)}{t}, \tag{7}$$

for any $t > 0$ if $\mathbf{E}(X)$ is finite. If the *kth moment* of X, $\mathbf{E}(X^k)$ is finite ($k > 0$), then one can apply (7) to $|X|^k$ getting

$$\mathbf{P}\left\{|X| \geq t\right\} = \mathbf{P}\left\{|X|^k \geq t^k\right\} \leq \frac{\mathbf{E}(|X|^k)}{t^k},$$

for any $t > 0$.

One can even get an upper bound going to 0 exponentially fast as $t \to \infty$ if $\mathbf{E}(e^{uX})$, the *moment generating function of* X, is finite for some $u_0 > 0$. For then, by (7),

$$\mathbf{P}\left\{X \geq t\right\} = \mathbf{P}\left\{u_0 X \geq u_0 t\right\} = \mathbf{P}\left\{e^{u_0 X} \geq e^{u_0 t}\right\} \leq e^{-u_0 t}\mathbf{E}(e^{u_0 X}), \tag{8}$$

for any $t > 0$.

Analogously, if $\mathbf{E}(e^{-u_0 X})$ is finite for some $u_0 > 0$, then

$$\mathbf{P}\{X \le -t\} = \mathbf{P}\{-u_0 X \ge u_0 t\} = \mathbf{P}\left\{e^{-u_0 X} \ge e^{u_0 t}\right\} \le e^{-u_0 t}\mathbf{E}(e^{-u_0 X}), \tag{9}$$

for any $t > 0$. Combining (8) and (9), one gets

$$\mathbf{P}\{|X| \ge t\} = \mathbf{P}\{X \ge t\} + \mathbf{P}\{X \le -t\} \le e^{-u_0 t}\left(\mathbf{E}(e^{u_0 X}) + \mathbf{E}(e^{-u_0 X})\right), \tag{10}$$

for any $t > 0$ if the moment generating function is finite both at u_0 and at $-u_0$.

Now, it is easy to find the moment generating function of one step of the random walk:

$$\mathbf{E}(e^{uX_k}) = e^u(1/2) + e^{-u}(1/2) = \cosh u.$$

Hence, using the independence of the steps, one obtains the moment generating function of the random walk S_n as

$$\mathbf{E}(e^{uS_n}) = \mathbf{E}(e^{u\sum_{k=1}^n X_k}) = \mathbf{E}(\prod_{k=1}^n e^{uX_k}) = (\cosh u)^n \quad (-\infty < u < \infty,\ n \ge 0). \tag{11}$$

Since $\cosh u$ is an even function and $\cosh 1 < 2$, (10) implies that

$$\mathbf{P}\{|S_n| \ge t\} \le 2 \cdot 2^n e^{-t} \quad (t > 0,\ n \ge 0). \tag{12}$$

3 Waiting Times

In the sequel we need the distribution of the random time τ when a random walk first hits either the point $x = 2$ or -2:

$$\tau = \tau_1 = \min\{n : |S_n| = 2\}. \tag{13}$$

To find the probability distribution of τ, imagine the random walk as a sequence of pairs of steps. These (independent) pairs can be classified either as a “return”: $(1, -1)$ or $(-1, 1)$, or as a “change of magnitude 2”: $(1, 1)$ or $(-1, -1)$. Both cases have the same probability $1/2$.

Clearly, it has zero probability that τ is equal to an odd number. The event $\{\tau = 2j\}$ occurs exactly when $j - 1$ “returns” are followed by a “change of magnitude 2”. Because of the independence of the pairs of steps, $\mathbf{P}\{\tau = 2j\} = 1/2^j$. It means that $\tau = 2Y$, where Y has geometric distribution with parameter $p = 1/2$,

$$\mathbf{P}\{\tau = 2j\} = \mathbf{P}\{Y = j\} = 1/2^j \quad (j \ge 1). \tag{14}$$

Hence,

$$\mathbf{E}(\tau) = 2\mathbf{E}(Y) = 2(1/p) = 4, \quad \mathbf{Var}(\tau) = 2^2\mathbf{Var}(Y) = 2^2(1-p)/p^2 = 8. \tag{15}$$

An important consequence is that with probability 1, a random walk sooner or later hits 2 or -2:

$$\mathbf{P}\{\tau < \infty\} = \sum_{j=1}^{\infty}(1/2^j) = 1.$$

It is also quite obvious that

$$\mathbf{P}\{S(\tau) = 2\} = \mathbf{P}\{S(\tau) = -2\} = 1/2. \tag{16}$$

This follows from the symmetry of the random walk. If we reflect $S(t)$ to the time axis, the resulting process $S^*(t)$ is also a random walk. Its corresponding τ^* is equal to τ, and the event $\{S^*(\tau) = 2\}$ is the same as $\{S(\tau) = -2\}$. Since $S^*(t)$ is just the same sort of random walk as $S(t)$, we have $\mathbf{P}\{S^*(\tau) = 2\} = \mathbf{P}\{S(\tau) = 2\}$ as well.

Another way to show (16) is to use the fact that the waiting time τ has countable many possible values and for any specific value we have symmetry:

$$\begin{aligned}
\mathbf{P}\{S(\tau) = 2\} &= \sum_{j=1}^{\infty} \mathbf{P}\{S(2j) = 2 \mid \tau = 2j\} \, \mathbf{P}\{\tau = 2j\} \\
&= \sum_{j=1}^{\infty} \mathbf{P}\{A_{2j-2}, X_{2j} = X_{2j-1} = 1 \mid A_{2j-2}, X_{2j} = X_{2j-1}\} \, \mathbf{P}\{\tau = 2j\} \\
&= (1/2) \sum_{j=1}^{\infty} \mathbf{P}\{\tau = 2j\} = 1/2,
\end{aligned}$$

where A_{2j-2} denotes the event that each of the first $j-1$ pairs is a “return”, i.e. $A_{2j-2} = \{X_2 = -X_1, \ldots, X_{2j-2} = -X_{2j-3}\}$, $A_0 = \emptyset$.

We mention that (16) illustrates a consequence of the so-called optional sampling theorem too: $\mathbf{E}S(\tau)) = 2\mathbf{P}\{S(\tau) = 2\} + (-2)\mathbf{P}\{S(\tau) = -2\} = 0$, which is the same as the expectation of $S(t)$.

We also need the probability of the event that a random walk starting from the point $x = 1$ hits $x = 2$ before hitting $x = -2$. This is equal to the conditional probability $\mathbf{P}\{S(\tau) = 2 \mid X_1 = 1\}$. If $X_1 = 1$, then $X_2 = 1$ with probability $1/2$, and then $\tau = 2$ and $S(\tau) = 2$ as well: $\mathbf{P}\{S(\tau) = 2, \tau = 2 \mid X_1 = 1\} = 1/2$.

On the other hand, if $X_1 = 1$, then $\tau > 2$ if and only if $X_2 = -1$, with probability $1/2$. that is, at the second step the walk returns the origin and starts “from scratch”. Then by (16), it has probability $1/2$ that the random walk hits 2 sooner than -2: $\mathbf{P}\{S(\tau) = 2, \tau > 2 \mid X_1 = 1\} = 1/4$. Therefore

$$\begin{aligned}
\mathbf{P}\{S(\tau) = 2 \mid X_1 = 1\} &= \mathbf{P}\{S(\tau) = 2, \tau = 2 \mid X_1 = 1\} + \mathbf{P}\{S(\tau) = 2, \tau > 2 \mid X_1 = 1\} \\
&= (1/2) + (1/4) = 3/4. \qquad (17)
\end{aligned}$$

It also follows then that

$$\mathbf{P}\{S(\tau) = -2 \mid X_1 = 1\} = 1 - (3/4) = 1/4. \tag{18}$$

(16), (17), and (18) are special cases of ruin probabilities [2, Section XIV,2]. For example, it can be shown that the probability that a random walk hits the level $a > 0$ before hitting the level $-b < 0$ is $b/(a+b)$.

Extending definition (13) of τ, for $k = 1, 2, \ldots$ we recursively define

$$\tau_{k+1} = \min \{n : n > 0, |S(T_k + n) - S(T_k)| = 2\},$$

where

$$T_k = T(k) = \tau_1 + \tau_2 + \cdots + \tau_k. \tag{19}$$

Then each τ_k has the same distribution as $\tau = \tau_1$. For,

$$\begin{aligned}
&\mathbf{P}\{\tau_{k+1} = 2j \mid T_k = 2m\} \\
&\quad = \mathbf{P}\{\ \min\{n : n > 0, |S(2m+n) - S(2m)| = 2\}\ = 2j \mid T_k = 2m\} \\
&\quad = \mathbf{P}\{\ \min\{n : n > 0, |S(n)| = 2\}\ = 2j\} = \mathbf{P}\{\tau_1 = 2j\} = 1/2^j,
\end{aligned}$$

where $k \geq 1$, $j \geq 1$, and $m \geq 1$ are arbitrary. The second equality above follows from two facts. First, each increment $S(2m+n) - S(2m)$ is independent of the event $\{T_k = 2m\}$, because the increment depends only on the random variables X_i $(2m+1 \leq i \leq 2m+n)$, while the event $\{T_k = 2m\}$ is determined exclusively by the random variables X_i $(1 \leq i \leq 2m)$, the corresponding "past". Second, each increment $S(2m+n) - S(2m)$ has the same distribution as $S(n)$, since both of them is a sum of n independent X_i. Hence, τ_{k+1} is independent of T_k (and also of any $\tau_i, i \leq k$), so indeed, $\mathbf{P}\{\tau_{k+1} = 2j\} = 1/2^j \quad (j \geq 1)$.

We also need the distribution of the random time T_k required by k changes of magnitude 2 along the random walk. In other words, $S(t)$ hits even integers (different from the previous one) exclusively at the time instants $T_1, T_2, \ldots$. To find the probability distribution of T_k, imagine the random walk again as a sequence of independent pairs of steps, "returns" and "changes of magnitude 2", both types having probability 1/2. The number of cases the event $\{T_k = 2j\}$ $(j \geq k)$ can occur is equal to the number of choices of $k-1$ pairs out of $j-1$ where a change of magnitude 2 occurs, before the last pair, which is necessarily a change of magnitude 2. Therefore

$$\mathbf{P}\{T_k = 2j\} = \binom{j-1}{k-1} \frac{1}{2^j} \quad (j \geq k \geq 1). \tag{20}$$

It means that $T_k = 2N_k$, where N_k has a negative binomial distribution with $p = 1/2$, [2, Section VI,8].

All this also follows from the fact that $N_k = T_k/2$ is the sum of k independent, geometrically distributed random variables with parameter $p = 1/2$, see (14) and (19): $N_k = Y_1 + Y_2 + \cdots + Y_k$ $(Y_j = \tau_j/2)$. Then T_k is finite valued with probability 1 and the expectation and variance of T_k easily follows from (15) and (19):

$$\mathbf{E}(T_k) = k\mathbf{E}(\tau) = 4k, \quad \mathbf{Var}(T_k) = k\mathbf{Var}(\tau) = 8k. \tag{21}$$

It is worth mentioning that T_k is a *stopping time* for each $k \geq 1$. By definition, it means that any event of the form $\{T_k \leq j\}$ depends exclusively on the corresponding "past" $S(t)$ $(t \leq j)$. In other words, $S_1, \ldots, S_j$ determine whether $\{T_k \leq j\}$ occurs or not.

Fortunately, the central limit and the large deviation theorems (see Theorem 1) can be proved for negative binomial distributions in the same fashion as for binomial distributions.

Theorem 2 *(a) For any real x fixed and $k \to \infty$ we have*

$$\mathbf{P}\left\{\frac{T_k - 4k}{\sqrt{8k}} \leq x\right\} \to \Phi(x).$$

(b) If $k \to \infty$ and $x_k \to \infty$ so that $x_k = o(k^{1/6})$, then

$$\mathbf{P}\left\{\frac{T_k - 4k}{\sqrt{8k}} \geq x_k\right\} \sim 1 - \Phi(x_k),$$

$$\mathbf{P}\left\{\frac{T_k - 4k}{\sqrt{8k}} \leq -x_k\right\} \sim \Phi(-x_k) = 1 - \Phi(x_k).$$

Proof. The normal approximation (4) is applicable to negative binomial distributions too: if $r = 2j$ and $k \to \infty$, then

$$\begin{aligned}
\mathbf{P}\{T_k = r\} &= \binom{j-1}{k-1}\frac{1}{2^j} = \frac{1}{2}\binom{j-1}{\frac{(j-1)+(2k-j-1)}{2}}\frac{1}{2^{j-1}} \\
&\sim \frac{1}{2}\frac{1}{\sqrt{\pi(j-1)/2}}\exp\left(-\frac{(\frac{2k-j-1}{2})^2}{(j-1)/2}\right) \\
&= \frac{1}{\sqrt{\pi(r-2)}}\exp\left(-\frac{(r-4k+2)^2}{4r-8}\right),
\end{aligned} \tag{22}$$

supposing $|2k - j - 1| = o((j-1)^{2/3})$, or equivalently,

$$|r - 4k| = o(k^{2/3}). \tag{23}$$

A routine computation shows that (22) is asymptotically equal to

$$\sim \frac{1}{\sqrt{4k\pi}}\exp\left(-\frac{(r-4k)^2}{16k}\right),$$

when $k \to \infty$ and (23) holds. Therefore we get an analogue of (4): if $k \to \infty$ and r is any even number such that $|r - 4k| < K_k = o(k^{2/3})$,

$$\mathbf{P}\{T_k = r\} \sim 2h\ \phi((r-4k)h), \ h = 1/\sqrt{8k}, \tag{24}$$

where ϕ denotes the standard normal density function.

Then in the same way as the statements of Theorem 1 are obtained from (4) in [2, Sections VII,3 and 6] one can get the present theorem from (24). Here we recall only the basic step of the argument:

$$\begin{aligned}\mathbf{P}\left\{x_1 \le \frac{T_k - 4k}{\sqrt{8k}} \le x_2\right\} \quad &\sim \sum_{\{r:\ x_1 \le (r-4k)h \le x_2,\ r \text{ is even}\}} 2h\ \phi((r-4k)h) \\ &\to \int_{x_1}^{x_2} \phi(t)\ dt = \Phi(x_2) - \Phi(x_1),\end{aligned}$$

for any x_1, x_2, when $k \to \infty$ and so $h \to 0$. The simple meaning of this is that Riemann sums converge to the corresponding integral. □

In the same fashion as the large deviation inequality (6) was obtained for S_n, Theorem 2(b) and (5) imply a large deviation type inequality for T_k:

$$\mathbf{P}\left\{\left|\frac{T_k - 4k}{\sqrt{8k}}\right| \ge x_k\right\} \le e^{-x_k^2/2}, \tag{25}$$

for $k \ge k_0$, supposing $x_k \to \infty$ and $x_k = o(k^{1/6})$ as $k \to \infty$.

Like in case of S_n, with T_k too we need a substitute for the large deviation inequality if the assumptions $k \to \infty$ or $x_k = o(k^{1/6})$ do not hold. The moment generating function of τ_n is simple:

$$\mathbf{E}(e^{u\tau_n}) = \sum_{j=1}^{\infty} e^{u2j} \frac{1}{2^j} = \frac{e^{2u}/2}{1 - (e^{2u}/2)} = \frac{1}{2e^{-2u} - 1}. \tag{26}$$

This function is finite if $u < \log\sqrt{2}$. Here and afterwards log denotes logarithm with base e.

Now the moment generating function of T_k follows from the independence of the τ_n's as

$$\mathbf{E}(e^{uT_k}) = \mathbf{E}(e^{u\sum_{n=1}^k \tau_n}) = \mathbf{E}(\prod_{n=1}^{k} e^{u\tau_n}) = (2e^{-2u} - 1)^{-k} \quad (u < \log\sqrt{2},\ k \ge 0). \tag{27}$$

We also need the moment generating function of the centered and "normalized" random variable $(T_k - 4k)/\sqrt{8}$, whose expectation is 0 and variance is k:

$$\mathbf{E}(e^{u(T_k - 4k)/\sqrt{8}}) = e^{-4ku/\sqrt{8}}\ \mathbf{E}(e^{T_k u/\sqrt{8}}) = \left(2e^{u/\sqrt{2}} - e^{u\sqrt{2}}\right)^{-k}, \tag{28}$$

for $u < \sqrt{2}\log 2$ and $k \ge 0$. Since (28) is less than 2^k for $u = \pm 1/2$, the exponential Chebyshev's inequality (10) implies that

$$\mathbf{P}\left\{|T_k - 4k|/\sqrt{8} \ge t\right\} \le 2 \cdot 2^k e^{-t/2} \quad (t > 0,\ k \ge 0). \tag{29}$$

4 From Random Walks to the Wiener Process: "Twist and Shrink"

Our construction of the Wiener process is based on P. Révész's one, [6, Section 6.2], which in turn is a simpler version of F.B. Knight's [4, Section 1.3]. The advantage of this method over the several known ones is that it is very natural and elementary.

We will define a sequence of approximations to the Wiener process, each of which is a "twisted and shrunk" random walk, a refinement of the previous one. It will be shown that this sequence converges to a process having the properties characterizing the Wiener process.

Imagine that we observe a particle undergoing Brownian motion. In the first experiment we observe the particle exclusively when it hits points with integer coordinates $j \in \mathbf{Z}$. Suppose that it happens exactly at the consecutive time instants $1, 2, \ldots$. To model the graph of the particle between the vertices so obtained the simplest idea is to join them by straight line segments like in Figure 1. Therefore the first approximation is

$$B_0(t) = S_0(t) = S(t),$$

where $t \geq 0$ real and $S(t)$ is a random walk defined by (1) and (2).

Suppose that in the second experiment we observe the particle when it hits points with coordinates $j/2$ ($j \in \mathbf{Z}$), in the third experiment when it hits points with coordinates $j/2^2$ ($j \in \mathbf{Z}$), etc. To model the second experiment one idea is to take a second random walk $S_1(t)$, independent of the first one, and shrink it.

Then the first problem that arises is the relationship between the time and space scales: if one wants to compress the length of a step into half, how much one has to compress the time needed for one step to preserve the essential properties of a random walk. Here we recall that by (3), the square root of the average squared distance of the random walk from the origin after time n is $\sqrt{n}$. So shrinking the random walk so that there are n steps in one time unit, each step should have a length $1/\sqrt{n}$. This way after one time unit the square root of the average squared distance of the walk from the origin will be around one spatial unit, like in the case of the original random walk. It means that compressing the length of one step into $1/2$ (or in general: $1/2^m, m = 1, 2, \ldots$) one has to compress the time needed for one step into $1/2^2$ (in general: $1/2^{2m}$).

The second problem is that sample-functions of $B_0(t)$ and of a shrunk version of an independent $S_1(t)$ have nothing to do with each other, the second is not being a refinement of the first in general. For example, if $B_0(1) = 1$, then it is equally likely that the first integer the shrunk version of $S_1(t)$ hits is $+1$ or -1.

Hence before shrinking we want to modify $S_1(t)$ so that it hits even integers $2j$ ($j \in \mathbf{Z}$) (counting the next one only if it is different from the previous one) in exactly the same order as $S_0(t)$ hits the corresponding integers $j \in \mathbf{Z}$. For example, if $S_0(1) = 1$ and $S_0(2) = 2$, then the first even integer $S_1(t)$ hits should be 2 and the next one (different from 2) should be 4. Thus if $S_1(t)$ hits the first even integer at time $T_1(1)$ and $S_1(T_1(1))$ happens to be -2, we will reflect every step $X_1(k)$ of $S_1(t)$ for $0 < k \leq T_1(1)$. This way we get a modified random walk $\tilde{S}_1(t)$ up to time $T_1(1)$ so that $\tilde{S}_1(T_1(1)) = 2$. Then we continue similarly up to

time $T_1(2)$: if the (already modified) walk hit 0 at time $T_1(2)$ (instead of 4), then we would reflect the steps $X_1(k)$ for $T_1(1) < k \leq T_1(2)$. This modification process, which we will call "twisting", ensures that the next approximation will always be a refinement of the previous one.

Now let us see the construction in detail. It begins with a sequence of independent random walks $S_0(t), S_1(t), \ldots$. That is, for each $m \geq 0$,

$$S_m(0) = 0, \; S_m(n) = X_m(1) + X_m(2) + \cdots + X_m(n) \; (n \geq 1), \tag{30}$$

where $X_m(k)$ $(m \geq 0, k \geq 1)$ is a double array of independent, identically distributed random variables such that

$$\mathbf{P}\{X_m(k) = 1\} = \mathbf{P}\{X_m(k) = -1\} = 1/2. \tag{31}$$

First we possibly modify $S_1(t), S_2(t), \ldots$ one-by-one, using the "twist" method to obtain a sequence of *not* independent random walks $\tilde{S}_1(t), \tilde{S}_2(t), \ldots$, each of which is a refinement of the former one. Second, by shrinking we get a sequence $B_1(t), B_2(t), \ldots$ approximating the Wiener process.

In accordance with the notation in (19), for $m \geq 1$, S_m hits even integers (different from the previous one) exclusively at the random time instants

$$T_m(0) = 0, \; T_m(k) = \tau_m(1) + \tau_m(2) + \cdots + \tau_m(k) \; (k \geq 1).$$

Each random variable $T_m(k)$ has the same distribution as $T(k) = T_k$ above, see (20) and (21). That is, $T_m(k)$ is the double of a negative binomial random variable, with expectation $4k$ and variance $8k$.

Now we define a suitable sequence of "twisted" random walks $\tilde{S}_m(t)$ $(m \geq 1)$ recursively, using $\tilde{S}_{m-1}(t)$, starting with

$$\tilde{S}_0(t) = S_0(t) \; (t \geq 0).$$

First we set

$$\tilde{S}_m(0) = 0.$$

Then for $k = 0, 1, \ldots$ successively and for every n such that $T_m(k) < n \leq T_m(k+1)$, we take (Figures 2-4).

$$\tilde{X}_m(n) = \begin{cases} X_m(n) & \text{if} \quad S_m(T_m(k+1)) - S_m(T_m(k)) = 2\tilde{X}_{m-1}(k+1); \\ -X_m(n) & \text{otherwise.} \end{cases} \tag{32}$$

and

$$\tilde{S}_m(n) = \tilde{S}_m(n-1) + \tilde{X}_m(n). \tag{33}$$

Observe that the stopping times $\tilde{T}_m(k)$ corresponding to $\tilde{S}_m(t)$ are the same as the original ones $T_m(k)$ $(m \geq 0, k \geq 0)$.

Lemma 1 *For each $m \geq 0$, $\tilde{S}_m(t)$ $(t \geq 0)$ is a random walk, that is, $\tilde{X}_m(1), \tilde{X}_m(2), \ldots$ is a sequence of independent, identically distributed random variables such that*

$$\mathbf{P}\left\{\tilde{X}_m(n) = 1\right\} = \mathbf{P}\left\{\tilde{X}_m(n) = -1\right\} = 1/2 \quad (n \geq 1). \tag{34}$$

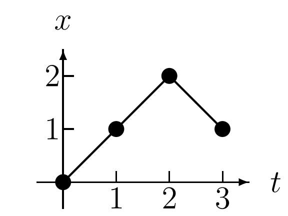

Figure 2: $B_0(t;\omega) = S_0(t;\omega)$.

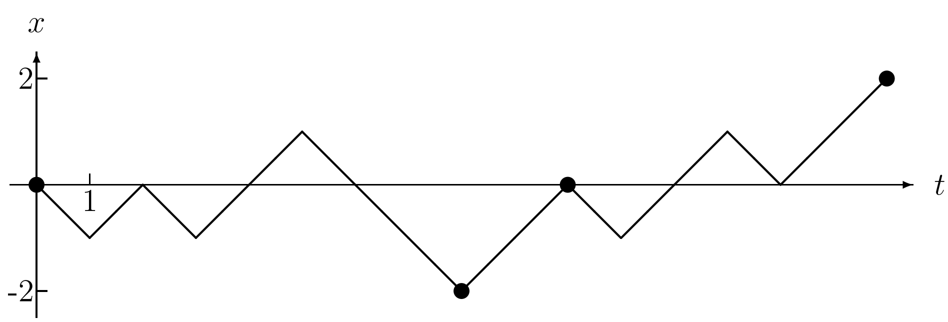

Figure 3: $S_1(t;\omega)$.

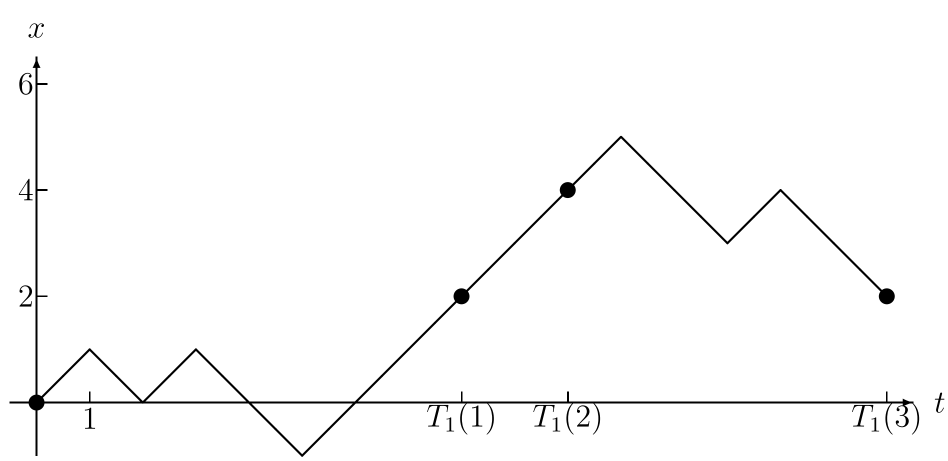

Figure 4: $\tilde{S}_1(t;\omega)$.

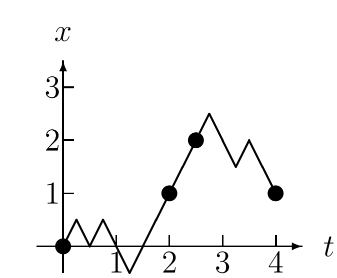

Figure 5: $B_1(t;\omega)$.

Proof. We proceed by induction over $m \geq 0$. For $m = 0$, $\tilde{S}_0(t) = S_0(t)$, a random walk by definition. So assume that $\tilde{S}_{m-1}(t)$ is a random walk, where $m \geq 1$, and see if it implies that $\tilde{S}_m(t)$ is a random walk too.

It is enough to show that for any $n \geq 1$ and any $\epsilon_j = \pm 1$ $(j = 1, \ldots, n)$ we have

$$\mathbf{P}\left\{\tilde{X}_m(1) = \epsilon_1, \ldots, \tilde{X}_m(n-1) = \epsilon_{n-1}, \tilde{X}_m(n) = \epsilon_n\right\} = 1/2^n. \tag{35}$$

Set the events $A_{m,r} = \left\{\tilde{X}_m(j) = \epsilon_j,\ 1 \leq j \leq r\right\}$ for $1 \leq r \leq n$ ($A_{m,0}$ is the sure event by definition) and the random variables $\Delta S^*_{m,k} = S_m(T_m(k+1)) - S_m(T_m(k))$ for $k \geq 0$. The event $A_{m,n-1}$ determines the greatest integer $k \geq 0$ such that $T_m(k) \leq n-1$; let us denote this value by κ. By (32),

$$\mathbf{P}\left\{A_{m,n}\right\} = \sum_{\alpha=\pm 1} \mathbf{P}\left\{A_{m,n-1}; X_m(n) = \alpha\epsilon_n;\ \Delta S^*_{m,\kappa} = \alpha 2\tilde{X}_{m-1}(\kappa+1)\right\}. \tag{36}$$

The event $A_{m,n-1}$ can be written here as $B_{m,n-1}C_{m,n-1}$, where

$$B_{m,n-1} = \left\{\tilde{X}_m(j) = \epsilon_j,\ 1 \leq j \leq T_m(\kappa)\right\},$$

$$C_{m,n-1} = \left\{X_m(j) = \alpha\epsilon_j,\ T_m(\kappa)+1 \leq j \leq n-1\right\}.$$

Definition (32) shows that $B_{m,n-1}$ is determined by $\tilde{X}_{m-1}(j)$ $(1 \leq j \leq \kappa)$ and $X_m(j)$ $(1 \leq j \leq T_m(\kappa))$ the values of which do not influence anything else in (36).

Then we distinguish two cases according to the parity of n.

Case 1: *n is odd.* Then $n-1$ is even and $S_m(T_m(\kappa)) = S_m(n-1)$. Further, let $\tau_{m,r} = \min\left\{j : j > 0, |S_m(r+j) - S_m(r)| = 2\right\}$ and $\Delta S_m(r) = S_m(r + \tau_{m,r}) - S_m(r)$ for $r \geq 0$. Then $S_m(T_m(\kappa+1)) = S_m(n-1+\tau_{m,n-1})$ and $\Delta S^*_{m,\kappa} = \Delta S_m(n-1)$. These and the argument above shows that in (36) $A_{m,n-1}$ is independent of the other terms. Consequently, (36) simplifies as

$$\mathbf{P}\left\{A_{m,n}\right\} = 2\mathbf{P}\left\{A_{m,n-1}\right\} \frac{1}{2} \sum_{\beta=\pm 1} \mathbf{P}\left\{X_m(n) = \epsilon_n;\ \Delta S_m(n-1) = 2\beta\right\}, \tag{37}$$

since the value of α is immaterial and $\mathbf{P}\left\{\tilde{X}_{m-1}(\kappa+1) = \beta\right\} = 1/2$, independently of everything else here.

Finally, (17) and (18) can be applied to (37):

$$\begin{aligned}
\mathbf{P}\left\{A_{m,n}\right\} &= \mathbf{P}\left\{A_{m,n-1}\right\} \sum_{\beta=\pm 1} \mathbf{P}\left\{\Delta S_m(n-1) = 2\beta \mid X_m(n) = \epsilon_n\right\}\ \mathbf{P}\left\{X_m(n) = \epsilon_n\right\} \\
&= \mathbf{P}\left\{A_{m,n-1}\right\}\left(\frac{3}{4} + \frac{1}{4}\right)\frac{1}{2} = \frac{1}{2}\mathbf{P}\left\{A_{m,n-1}\right\},
\end{aligned}$$

independently of ϵ_n.

CASE 2: *n is even.* Then $n-2$ is even and the argument in Case 1 could be repeated with $n-2$ in place of $n-1$, with the only exception that in (36) we have an additional term $\tilde{X}_m(n-1) = \alpha X_m(n-1)$. Then instead of (37) we arrive at

$$\begin{aligned}
&\mathbf{P}\{A_{m,n}\} \\
&= \mathbf{P}\{A_{m,n-2}\} \sum_{\beta=\pm 1} \mathbf{P}\{X_m(n-1)=\epsilon_{n-1}, X_m(n)=\epsilon_n;\ \Delta S_m(n-2)=2\beta\} \\
&= \mathbf{P}\{A_{m,n-2}\} \frac{1}{2^2} \sum_{\beta=\pm 1} \mathbf{P}\{\Delta S_m(n-2)=2\beta \mid X_m(n-1)=\epsilon_{n-1}, X_m(n)=\epsilon_n\}. \quad (38)
\end{aligned}$$

The conditional probability in (38) is

$$\begin{array}{llllll}
1 & \text{if} & \beta=\epsilon_{n-1}=\epsilon_n; & 0 & \text{if} & -\beta=\epsilon_{n-1}=\epsilon_n; \\
1/2 & \text{if} & \beta=\epsilon_{n-1}=-\epsilon_n; & 1/2 & \text{if} & -\beta=\epsilon_{n-1}=-\epsilon_n.
\end{array}$$

Thus the sum in (38) becomes

$$1+0=1 \text{ if } \quad \epsilon_{n-1}=\epsilon_n; \qquad (1/2)+(1/2)=1 \text{ if } \quad \epsilon_{n-1}=-\epsilon_n.$$

In other words, the value of the sum in (38) is 1, independently of ϵ_{n-1} and ϵ_n.

In summary, $\mathbf{P}\{A_{m,n}\} = \frac{1}{2}\mathbf{P}\{A_{m,n-1}\}$ if n is odd and $\mathbf{P}\{A_{m,n}\} = \frac{1}{4}\mathbf{P}\{A_{m,n-2}\}$ if n is even. Since $\mathbf{P}\{A_{m,0}\}=1$, (35) follows. □

We mention that an other possibility to prove Lemma 1 is to introduce the random variables $Z_k = \frac{1}{2}\Delta S^*_{m,k-1}\tilde{X}_{m-1}(k)$ for $k \geq 1$. It can be shown that $Z_1, Z_2, \ldots$ is a sequence of independent and identically distributed random variables, $\mathbf{P}\{Z_k=1\} = \mathbf{P}\{Z_k=-1\} = 1/2$, and this sequence is independent of the sequence $X_m(1), X_m(2), \ldots$ as well. Then we have $\tilde{X}_m(n) = Z_k X_m(n)$ for each n such that $T_m(k-1) < n \leq T_m(k)$ $(k \geq 1)$ and this implies (35).

The main property that was aimed when we introduced the "twist" method easily follows from (32) and (33):

$$\tilde{S}_m(T_m(k)) = \sum_{j=1}^{k} \tilde{S}_m(T_m(j)) - \tilde{S}_m(T_m(j-1)) = \sum_{j=1}^{k} 2\tilde{X}_{m-1}(j) = 2\tilde{S}_{m-1}(k), \tag{39}$$

for any $m \geq 1$ and $k \geq 0$.

Now the second step of the approximation comes: "shrinking". As was discussed above, at the mth approximation the length of one step should be $1/2^m$ and the time needed for a step should be $1/2^{2m}$ (Figure 5). So we define the *mth approximation of the Wiener process* by

$$B_m\left(\frac{t}{2^{2m}}\right) = \frac{1}{2^m}\tilde{S}_m(t) \quad (t \geq 0, m \geq 0), \tag{40}$$

or equivalently, $B_m(t) = 2^{-m}\tilde{S}_m(t2^{2m})$. Basically, $B_m(t)$ is a model of Brownian motion on the set of points $x = j/2^m$ $(j \in \mathbf{Z})$.

Now (39) becomes the following *refinement property*:

$$B_m\left(\frac{T_m(k)}{2^{2m}}\right) = B_{m-1}\left(\frac{k}{2^{2(m-1)}}\right), \tag{41}$$

for any $m \geq 1$ and $k \geq 0$.

The remaining part of this section is devoted to showing the convergence of the sequence $B_m(t)$ $(m = 0, 1, 2, \ldots)$, and that the limiting process has the characterizing properties of the Wiener process. In proving these our basic tools will be some relatively simple, but powerful observations.

First, often in the sequel the following crude, but still efficient estimate will be applied:

$$\mathbf{P}\left\{\max_{1\leq j\leq N} Z_j \geq t\right\} = \mathbf{P}\left\{\bigcup_{j=1}^{N} \{Z_j \geq t\}\right\} \leq \sum_{j=1}^{N} \mathbf{P}\{Z_j \geq t\}, \tag{42}$$

which is valid for arbitrary random variables Z_j and real number t.

The proofs of Lemmas 3 and 4 below essentially consist of the application of the following large deviation type estimate fulfilled by S_n and $(T_k - 4k)/\sqrt{8}$ according to Theorems 1(b) and 2(b). The previously mentioned exponential Chebyshev's inequalities (12) and (29) will be also used. Note that in the next lemma we have $a = 2$ and $b = 1$ for S_n in (12) and $a = 2$ and $b = 1/2$ for $(T_k - 4k)/\sqrt{8}$ in (29).

Lemma 2 *Suppose that for $j \geq 0$, we have $\mathbf{E}(Z_j) = 0$, $\mathbf{Var}(Z_j) = j$, and with some $a > 0$ and $b > 0$,*

$$\mathbf{P}\{|Z_j| \geq t\} \leq 2a^j e^{-bt} \quad (t > 0)$$

(exponential Chebyshev-type inequality).

Assume as well that there exists a $j_0 > 0$ such that for any $j \geq j_0$,

$$\mathbf{P}\left\{|Z_j|/\sqrt{j} \geq x_j\right\} \leq e^{-x_j^2/2},$$

whenever $x_j \to \infty$ and $x_j = o(j^{1/6})$ as $j \to \infty$ (large deviation type inequality).

Then for any $C > 1$,

$$\mathbf{P}\left\{\max_{0\leq j\leq N} |Z_j| \geq \sqrt{2CN\log N}\right\} \leq \frac{2}{N^{C-1}}, . \tag{43}$$

if N is large enough, $N \geq N_0(C)$.

Proof. The maximum in (43) can be handled by the crude estimate (42). Divide the resulting sum into two parts: one that can be estimated by a large deviation type inequality, and an other that will be estimated using exponential Chebyshev's inequality. For the large deviation part x_j will be $\sqrt{2C\log N}$. Since $j \leq N$, $j \to \infty$ implies that $N \to \infty$, and then

$x_j \to \infty$ as well. If $j \geq \log^4 N$, then the condition $x_j = o(j^{1/6})$ holds too, and the large deviation type inequality is applicable. Thus

$$
\begin{aligned}
& \mathbf{P}\left\{\max_{0 \leq j \leq N} |Z_j| \geq \sqrt{2CN \log N}\right\} \\
\leq & \sum_{j=0}^{\lfloor \log^4 N \rfloor} 2a^j \exp\left(-b\sqrt{2CN\log N}\right) + \sum_{j=\lfloor \log^4 N \rfloor}^{N} \mathbf{P}\left\{|Z_j|/\sqrt{j} \geq \sqrt{2C \log N}\right\} \\
\leq & \frac{2a}{a-1} \exp\left(\log a \ \log^4 N - b\sqrt{2CN \log N}\right) + N \exp(-C \log N) \leq 2 \ N^{1-C}
\end{aligned}
$$

if $C > 1$ and $N \geq N_0(C)$. ($\lfloor x \rfloor$ denotes the greatest integer less than or equal to x.) □

Note that the lemma and its proof are valid even when N is not an integer. Here and afterwards we use the convention that if the upper limit of a sum is a real value N, then the sum goes until $\lfloor N \rfloor$.

We mention that both inequalities among the assumptions of the previous Lemma 2 hold for partial sums Z_j of any sequence of independent and identically distributed random variables with expectation 0, variance 1 and a moment generating function which is finite for some $\pm u_0$. The fact that an exponential Chebyshev-type inequality should hold then can be seen from (10) and (11), while the large deviation type estimate is shown to hold e.g. in [3, Section XVI,6].

The *first Borel–Cantelli lemma* [2, Section VIII,3] is also an important tool, stating that if there is given an infinite sequence $A_1, A_2, \dots$ of events such that $\sum_{m=1}^{\infty} \mathbf{P}\{A_m\}$ is finite, then *with probability* 1 only finitely many of the events occur. Or with an other widely used terminology: *almost surely* only finitely many of them will occur.

Now turning to the convergence proof, as the first step, it will be shown that the time instants $T_{m+1}(k)/2^{2(m+1)}$ will get arbitrarily close to the time instants $k/2^{2m} = 4k/2^{2(m+1)}$ as $m \to \infty$. By (41), this means that the next approximation not only visits the points $x = j/2^m$ ($j \in \mathbf{Z}$) in the same order, but the corresponding time instants will get arbitrarily close to each other as $m \to \infty$. Remember that by (20) and (21), $T_m(k)$ is the double of a negative binomial random variable, with expectation $4k$ and variance $8k$. Here Lemma 2 will be applied to $(T_m(k) - 4k)/\sqrt{8}$ with $N = K2^{2m}$. So $\log N = \log K + (2 \log 2)m \leq 1.5m$ if m is large enough, $m \geq m_0(K)$, and then $\sqrt{2CN \log N} \leq \sqrt{3CKm}\, 2^m$.

Lemma 3 *(a) For any $C > 1$, $K > 0$, and for any $m \geq m_0(C, K)$ we have*

$$
\mathbf{P}\left\{\max_{0 \leq k/2^{2m} \leq K} |T_{m+1}(k) - 4k| \geq \sqrt{24CKm}\, 2^m\right\} < 2\,(K2^{2m})^{1-C}. \tag{44}
$$

(b) For any $K > 0$,

$$
\max_{0 \leq k/2^{2m} \leq K} \left| \frac{T_{m+1}(k)}{2^{2(m+1)}} - \frac{k}{2^{2m}} \right| < \sqrt{2Km}\, 2^{-m} \tag{45}
$$

with probability 1 *for all but finitely many* m.

Proof.
(a) (44) is a direct consequence of Lemma 2.
(b) Take for example $C = 4/3$ in (a) and define the following events for $m \geq 0$:

$$A_m = \left\{ \max_{0 \leq k/2^{2m} \leq K} |T_{m+1}(k) - 4k| \geq \sqrt{32Km}\, 2^m \right\}.$$

By (44), for $m \geq m_0(C, K)$, $\mathbf{P}\{A_m\} < 2\left(K2^{2m}\right)^{-1/3}$. Then $\sum_{m=0}^{\infty} \mathbf{P}\{A_m\} < \infty$. Hence the Borel–Cantelli lemma implies that with probability 1, only finitely many of the events A_m occur. That is, almost surely for all but finitely many m we have

$$\max_{0 \leq k/2^{2m} \leq K} |T_{m+1}(k) - 4k| < \sqrt{32Km}\, 2^m.$$

This inequality is equivalent to (45). □

It seems to be important to emphasize a "weakness" of a statement like the one in Lemma 3(b): we use the phrase "all but finitely many m" to indicate that the statement holds for every $m \geq m_0(\omega)$, where $m_0(\omega)$ may depend on the specific point ω of the sample space. In other words, one has no common, uniform lower bound for m in general.

Next we want to show that for any $j \geq 1$, $B_{n+j}(t)$ will be arbitrarily close to $B_n(t)$ as $n \to \infty$. Here again Lemma 2 will be applied, this time to a random walk S_r, with a properly chosen N' and C' (instead of $N = K2^{2m}$ and C). Although the proof will be somewhat long, its basic idea is simple. Since $B_{m+1}(T_{m+1}(k)/2^{2(m+1)}) = B_m(k/2^{2m})$ by (39), and the difference of the corresponding time instants here approaches zero fast as $m \to \infty$ by (45), one can show that $B_m(t)$ and its refinement $B_{m+1}(t)$ will get very close to each other too.

The following elementary fact that we need in the proof is discussed before stating the lemma:

$$\sum_{m=n}^{\infty} m2^{-m/2} = (1/\sqrt{2}) \sum_{m=n}^{\infty} m \left(1/\sqrt{2}\right)^{m-1} < 4n2^{-n/2}, \tag{46}$$

for $n \geq 15$. This can be shown by a routine application of power series:

$$\sum_{m=n}^{\infty} mx^{m-1} = \frac{d}{dx} \sum_{m=n}^{\infty} x^m = \frac{d}{dx}\left(\frac{x^n}{1-x}\right) = nx^{n-1}\left(\frac{1}{1-x} + \frac{x}{n(1-x)^2}\right).$$

Substituting $x = 1/\sqrt{2}$, one gets (46) for $n \geq 15$.

Lemma 4 *(a) For any $C \geq 3/2$, $K > 0$, and for any $n \geq n_0(C, K)$ we have*

$$\mathbf{P}\left\{ \max_{0 \leq k/2^{2n} \leq K} |B_{n+1}(T_{n+1}(k)/2^{2(n+1)}) - B_{n+1}(k/2^{2n})| \geq (1/8)n2^{-n/2} \right\} \leq 3(K2^{2n})^{1-C} \tag{47}$$

and

$$\mathbf{P}\left\{ \max_{0 \leq t \leq K} |B_{n+j}(t) - B_n(t)| \geq n2^{-n/2} \;\; \textit{for some } j \geq 1 \right\} < 6(K2^{2n})^{1-C}. \tag{48}$$

(b) For any $K > 0$,

$$\max_{0\le t\le K} |B_{n+j}(t) - B_n(t)| < n2^{-n/2}, \tag{49}$$

with probability 1 for all $j \ge 1$ and for all but finitely many n.

PROOF. Let us consider first the difference between two consecutive approximations, $\max_{0\le t\le K} |B_{m+1}(t) - B_m(t)|$. The maximum over real values t can be approximated by the maximum over dyadic rational numbers $k/2^{2m}$. This is true because any sample-function $B_m(t;\omega)$ is a broken line such that, by (40), the magnitude of the increment between two consecutive points $k/2^{2m}$ and $(k+1)/2^{2m}$ is equal to 2^{-m}. Thus, taking the integer $t_m = \lfloor t2^{2m} \rfloor$ for each $t \in [0, K]$, one has $t_m/2^{2m} \le t < (t_m+1)/2^{2m}$ and so $4t_m/2^{2(m+1)} \le t < (4t_m+4)/2^{2(m+1)}$. So we get $|B_m(t) - B_m(t_m/2^{2m})| < 2^{-m}$ and $\left|B_{m+1}(t) - B_{m+1}\left(4t_m/2^{2(m+1)}\right)\right| \le 4 \cdot 2^{-(m+1)} = 2 \cdot 2^{-m}$. Hence

$$\max_{0\le t\le K} |B_{m+1}(t) - B_m(t)| \le \max_{0\le k/2^{2m}\le K} \left|B_{m+1}\left(4k/2^{2(m+1)}\right) - B_m\left(k/2^{2m}\right)\right| + 3 \cdot 2^{-m}.$$

Moreover, by (41) and (40) we have

$$\begin{aligned} B_{m+1}\left(4k/2^{2(m+1)}\right) - B_m\left(k/2^{2m}\right) &= B_{m+1}\left(4k/2^{2(m+1)}\right) - B_{m+1}\left(T_{m+1}(k)/2^{2(m+1)}\right) \\ &= 2^{-(m+1)}\tilde{S}_{m+1}(4k) - 2^{-(m+1)}\tilde{S}_{m+1}(T_{m+1}(k)). \end{aligned} \tag{50}$$

Thus

$$\begin{aligned} &\mathbf{P}\left\{\max_{0\le t\le K} |B_{m+1}(t) - B_m(t)| \ge (1/4)m2^{-m/2}\right\} \\ &\le \mathbf{P}\left\{\max_{0\le k/2^{2m}\le K} \left|B_{m+1}\left(4k/2^{2(m+1)}\right) - B_m\left(k/2^{2m}\right)\right| \ge (1/8)m2^{-m/2}\right\} \\ &= \mathbf{P}\left\{\max_{0\le k\le K2^{2m}} |\tilde{S}_{m+1}(4k) - \tilde{S}_{m+1}(T_{m+1}(k))| \ge (1/4)m2^{m/2}\right\} \end{aligned} \tag{51}$$

if m is large enough.

By Lemma 3, the probability of the event

$$A_m = \left\{\max_{0\le k\le K2^{2m}} |T_{m+1}(k) - 4k| \ge \sqrt{24CKm}\, 2^m\right\}$$

is very small for m large. Therefore divide the last expression in (51) into two parts according to A_m and A_m^c (the complement of A_m):

$$\begin{aligned} &\mathbf{P}\left\{\max_{0\le t\le K} |B_{m+1}(t) - B_m(t)| \ge (1/4)m2^{-m/2}\right\} \\ &\le \mathbf{P}\left\{A_m^c; \max_{0\le k\le K2^{2m}} \left|\tilde{S}_{m+1}(4k) - \tilde{S}_{m+1}(T_{m+1}(k))\right| \ge (1/4)m2^{m/2}\right\} + \mathbf{P}\left\{A_m\right\} \\ &\le \sum_{k=1}^{K2^{2m}} \mathbf{P}\left\{\max_{\left\{j:\ |j-4k|<\sqrt{24CKm}\, 2^m\right\}} \left|\tilde{S}_{m+1}(4k) - \tilde{S}_{m+1}(j)\right| \ge (1/4)m2^{m/2}\right\} \\ &\quad +2(K2^{2m})^{1-C}, \end{aligned} \tag{52}$$

where the crude estimate (42) and Lemma 3(a) were used.

Now apply Lemma 2 to $\tilde{S}_{m+1}(j) - \tilde{S}_{m+1}(4k)$ here, with suitably chosen N' and C'. For k fixed and $j > 4k$, $\tilde{S}_{m+1}(j) - \tilde{S}_{m+1}(4k) = \sum_{r=4k+1}^{j} X_{m+1}(r)$ is a random walk of the form $S(j-4k)$. (The case $j < 4k$ is symmetric.) Since $|j-4k| < \sqrt{24CKm}\, 2^m$, N' is taken as $\sqrt{24CKm}\, 2^m$. (So N' is roughly $\sqrt{N}$, where $N = K2^{2m}$.) Then $\log N' = (1/2)\log(24CKm)$ $+(\log 2)m \le m$ if m is large enough, depending on C and K. So

$$\sqrt{2C'N'\log N'} \le \sqrt{2C'm\sqrt{24CKm}\, 2^m} \le (1/4)m2^{m/2},$$

if m is large enough, depending on C, C', and K. Then it follows by Lemma 2 that

$$\mathbf{P}\left\{ \max_{0 \le r \le \sqrt{24CKm}\, 2^m} |S(r)| \ge (1/4)m2^{m/2} \right\} \le 2(\sqrt{24CKm}\, 2^m)^{1-C'}. \tag{53}$$

The second term of the error probability in (52) is $2(K2^{2m})^{1-C} = 2N^{1-C}$, while (53) indicates that the first term is at most $K2^{2m} \cdot 2 \cdot 2(\sqrt{24CKm}\, 2^m)^{1-C'} \le N(\sqrt{N})^{1-C'}$ if $C' > 1$ and m is large enough. To make the two error terms to be of the same order, choose $1 + (1-C')/2 = 1 - C$, i.e. $C' = 2C + 1$. Thus (52) becomes

$$\mathbf{P}\left\{ \max_{0 \le t \le K} |B_{m+1}(t) - B_m(t)| \ge (1/4)m2^{-m/2} \right\} \le 3(K2^{2m})^{1-C},$$

for any m large enough, depending on C and K. Comparing this to (50) and (51) one obtains (47).

By (46), $\max_{0 \le t \le K} |B_{m+1}(t) - B_m(t)| < (1/4)m2^{-m/2}$ for all $m \ge n \ge 15$ would imply that

$$\begin{aligned} \max_{0 \le t \le K} |B_{n+j}(t) - B_n(t)| &= \max_{0 \le t \le K} \left| \sum_{m=n}^{n+j-1} B_{m+1}(t) - B_m(t) \right| \\ \le \sum_{m=n}^{n+j-1} \max_{0 \le t \le K} |B_{m+1}(t) - B_m(t)| &< \sum_{m=n}^{\infty} (1/4)m2^{-m/2} < n2^{-n/2}, \end{aligned}$$

for any $j \ge 1$. So we conclude that

$$\begin{aligned} &\mathbf{P}\left\{ \max_{0 \le t \le K} |B_{n+j}(t) - B_n(t)| \ge n2^{-n/2} \quad \text{for some } j \ge 1 \right\} \\ &\le \sum_{m=n}^{\infty} \mathbf{P}\left\{ \max_{0 \le t \le K} |B_{m+1}(t) - B_m(t)| \ge (1/4)m2^{-m/2} \right\} \\ &\le \sum_{m=n}^{\infty} 3(K2^{2m})^{1-C} = 3(K2^{2n})^{1-C} \frac{1}{1 - 2^{2(1-C)}} < 6(K2^{2n})^{1-C} \end{aligned}$$

if $C \ge 3/2$ (say), for any $n \ge n_0(C, K)$. This proves (48).

The statement in (b) follows from (48) by the Borel–Cantelli lemma, as in the proof of Lemma 3. $\square$

Now we are ready to establish the existence of the Wiener process, which is a continuous model of Brownian motion. An important consequence of (49) is that the difference between the Wiener process and $B_n(t)$ is smaller than a constant multiple of $\log N/\sqrt[4]{N}$, where $N = K2^{2n}$, see (55) below.

Theorem 3 *As* $n \to \infty$, *with probability* 1 *(that is, for almost all* $\omega \in \Omega$*) and for all* $t \in [0, \infty)$ *the sample-functions* $B_n(t; \omega)$ *converge to a sample-function* $W(t; \omega)$ *such that*

(i) $W(0; \omega) = 0$, $W(t; \omega)$ *is a continuous function of* t *on the interval* $[0, \infty)$*;*

(ii) for any $0 \le s < t$, $W(t) - W(s)$ *is a normally distributed random variable with expectation* 0 *and variance* $t - s$*;*

(iii) for any $0 \le s < t \le u < v$, *the increments* $W(t) - W(s)$ *and* $W(v) - W(u)$ *are independent random variables.*

By definition, $W(t)$ $(t \ge 0)$ *is called the* **Wiener process***.*

Further, we have the following estimates for the difference of the Wiener process and its approximations.

(a) For any $C \ge 3/2$, $K > 0$, *and for any* $n \ge n_0(C, K)$ *we have*

$$\mathbf{P}\left\{\max_{0 \le t \le K} |W(t) - B_n(t)| \ge n2^{-n/2}\right\} \le 6(K2^{2n})^{1-C}. \tag{54}$$

(b) For any $K > 0$,

$$\max_{0 \le t \le K} |W(t) - B_n(t)| \le n2^{-n/2}, \tag{55}$$

with probability 1 *for all but finitely many* n.

Proof. Lemma 4(b) shows that for almost all $\omega \in \Omega$, the sequence $B_n(t; \omega)$ converges for any $t \ge 0$ as $n \to \infty$. Let us denote the limit by $W(t; \omega)$. On a probability zero ω-set the limit possibly does not exist, there one can define $W(t; \omega) = 0$ for any $t \ge 0$. Since $B_n(0; \omega) = 0$ for any n, it follows that $W(0; \omega) = 0$ for any $\omega \in \Omega$.

Taking $j \to \infty$ in (48), (54) follows. By (49), the convergence of $B_n(t)$ is uniform on any bounded interval $[0, K]$, more exactly, for any $K > 0$ we have (55) with probability 1. Textbooks on advanced calculus, like W. Rudin's [7, Section 7.12] show that the limit function of a uniformly convergent sequence of continuous functions is also continuous. This proves (i).

Now we turn to the proof of (ii). Take arbitrary $t > s \ge 0$ and x real. With $K > t$ fixed, (54) shows that for any $\delta > 0$ there exists an $n \ge n_0(C, K)$ such that

$$\mathbf{P}\left\{\max_{0 \le u \le K} |W(u) - B_n(u)| \ge \delta\right\} < \delta. \tag{56}$$

Since

$$\mathbf{P}\{W(t) - W(s) \le x\} = \mathbf{P}\{B_n(t) - B_n(s) \le x - (W(t) - B_n(t)) + (W(s) - B_n(s))\},$$

(56) implies that

$$\begin{aligned} \mathbf{P}\{B_n(t) - B_n(s) \le x - 2\delta\} - 2\delta &\le \mathbf{P}\{W(t) - W(s) \le x\} \\ &\le \mathbf{P}\{B_n(t) - B_n(s) \le x + 2\delta\} + 2\delta. \end{aligned} \tag{57}$$

This indicates that the distribution function of $W(t) - W(s)$ can be eventually obtained from the distribution function of

$$B_n(t) - B_n(s) = 2^{-n}\tilde{S}_n(2^{2n}t) - 2^{-n}\tilde{S}_n(2^{2n}s). \tag{58}$$

Take the non-negative integers $j_n = \lfloor 2^{2n}t \rfloor$ and $i_n = \lfloor 2^{2n}s \rfloor$, $j_n \geq i_n$. Then (58) differs from

$$2^{-n}(\tilde{S}_n(j_n) - \tilde{S}_n(i_n)) = 2^{-n} \sum_{k=i_n+1}^{j_n} \tilde{X}_k \tag{59}$$

by an error not more than $2 \cdot 2^{-n} < \delta$. (We can assume that n was chosen so.) Also, $j_n - i_n$ differs from $2^{2n}(t-s)$ by at most 1. In particular, $j_n - i_n \to \infty$ as $n \to \infty$.

If n is large enough (we can assume again that n was chosen so), by Theorem 1(a), for any fixed real x' we have

$$\Phi(x') - \delta \leq \mathbf{P}\left\{ \frac{1}{\sqrt{j_n - i_n}} \sum_{k=i_n+1}^{j_n} \tilde{X}_k \leq x' \right\} \leq \Phi(x') + \delta. \tag{60}$$

Here $\sqrt{j_n - i_n}$ can be approximated by $2^n\sqrt{t-s}$ if n is large enough,

$$1 - \delta < \sqrt{\frac{j_n - i_n - 1}{j_n - i_n}} \leq \frac{2^n\sqrt{t-s}}{\sqrt{j_n - i_n}} \leq \sqrt{\frac{j_n - i_n + 1}{j_n - i_n}} < 1 + \delta. \tag{61}$$

Combining formulae (58)-(61) we obtain that

$$\Phi\left((1-\delta)\frac{x}{\sqrt{t-s}} - \delta\right) - \delta \leq \mathbf{P}\left\{B_n(t) - B_n(s) \leq x\right\} \leq \Phi\left((1+\delta)\frac{x}{\sqrt{t-s}} + \delta\right) + \delta.$$

This shows that the distribution of $B_n(t) - B_n(s)$ is asymptotically normal with mean 0 and variance $t - s$ as $n \to \infty$. Moreover, by (57), the distribution of $W(t) - W(s)$ is *exactly* normal with mean 0 and variance $t - s$, since δ can be made arbitrarily small if n is large enough:

$$\mathbf{P}\left\{W(t) - W(s) \leq x\right\} = \Phi\left(\frac{x}{\sqrt{t-s}}\right).$$

This proves (ii).

Finally, (iii) can be proved similarly as (ii) above. Taking arbitrary $v > u \geq t > s \geq 0$ and x, y real numbers,

$$\mathbf{P}\left\{W(t) - W(s) \leq x,\ W(v) - W(u) \leq y\right\} \tag{62}$$

can be approximated by a probability of the form

$$\mathbf{P}\left\{B_n(t) - B_n(s) \leq x\ ,\ B_n(v) - B_n(u) \leq y\right\}$$

arbitrarily well if n is large enough, just like in (57). In turn, like in (59), the latter can be estimated arbitrarily well by a probability of the form

$$\mathbf{P}\left\{\frac{1}{\sqrt{j_n - i_n}} \sum_{k=i_n+1}^{j_n} \tilde{X}_k \leq x' \ , \ \frac{1}{\sqrt{r_n - q_n}} \sum_{k=q_n+1}^{r_n} \tilde{X}_k \leq y'\right\}, \tag{63}$$

where $i_n = \lfloor 2^{2n}s \rfloor \leq j_n = \lfloor 2^{2n}t \rfloor \leq q_n = \lfloor 2^{2n}u \rfloor \leq r_n = \lfloor 2^{2n}v \rfloor$.

Since there are no common terms in the first and the second sum of (63), the two sums are independent. Thus (63) is equal to

$$\mathbf{P}\left\{\frac{1}{\sqrt{j_n - i_n}} \sum_{k=i_n+1}^{j_n} \tilde{X}_k \leq x'\right\} \cdot \mathbf{P}\left\{\frac{1}{\sqrt{r_n - q_n}} \sum_{k=q_n+1}^{r_n} \tilde{X}_k \leq y'\right\},$$

which can be made arbitrarily close to

$$\mathbf{P}\left\{W(t) - W(s) \leq x\right\} \cdot \mathbf{P}\left\{W(v) - W(u) \leq y\right\}. \tag{64}$$

Since errors in the approximations can be made arbitrarily small, (62) and (64) must agree for any real x and y. This proves (iii). $\square$

Note that properties (ii) and (iii) are often rephrased in the way that the Wiener-process is a Gaussian process with independent and stationary increments. It can be proved [4, Section 1.5] that properties (i), (ii), and (iii) characterize the Wiener process. In other words, any construction to the Wiener process gives essentially the same process that was constructed above.

5 From the Wiener Process to Random Walks

Now we are going to check whether the Wiener process as a model of Brownian motion has the properties described in the introduction of Section 4. Namely, we would want to find the sequence of shrunk random walks $B_m(k2^{-2m})$ in $W(t)$.

Let $s(1)$ be the first (random) time instant where the magnitude of the Wiener process is 1: $s(1) = \min\{s > 0 : |W(s)| = 1\}$. The continuity and increment characteristics of the Wiener process imply that $s(1)$ exists with probability 1. Clearly, each shrunk random walk $B_m(t)$ has the symmetry property that reflecting all its sample-functions to the time axis, one gets the same process. $W(t)$ as a limiting process of shrunk random walks inherits this feature. Therefore setting $X(1) = W(s(1))$, $\mathbf{P}\{X(1) = 1\} = \mathbf{P}\{X(1) = -1\} = 1/2$.

Inductively, starting with $s(0) = 0$, if $s(k-1)$ is given, define the random time instant

$$s(k) = \min\{s : s > s(k-1), |W(s) - W(s(k-1))| = 1\} \quad (k \geq 1).$$

As above, $s(k)$ exists with probability 1. Setting $X(k) = W(s(k)) - W(s(k-1))$, it is heuristically clear that $\mathbf{P}\{X(k) = 1\} = \mathbf{P}\{X(k) = -1\} = 1/2$, and $X(k)$ is independent of $X(1), X(2), \ldots, X(k-1)$.

This way one gets a random walk $S(k) = W(s(k)) = X(1) + X(2) + \cdots X(k)$ $(k \geq 0)$ from the Wiener process. Using a more technical phrase, by this method, based on *first passage times*, one can *imbed* a random walk into the Wiener process; it is a special case of the famous Skorohod imbedding, see e.g. [1, Section 13.3].

Quite similarly, one can imbed $B_m(k2^{-2m})$ into $W(t)$ for any $m \geq 0$ by setting $s_m(0) = 0$,

$$s_m(k) = \min\left\{s : s > s_m(k-1), |W(s) - W(s_m(k-1))| = 2^{-m}\right\} (k \geq 1), \tag{65}$$

and $B_m(k2^{-2m}) = W(s_m(k))$ $(k \geq 0)$.

However, instead of proving all necessary details about Skorohod imbedding briefly described above, we will define an other imbedding method better suited to our approach. It will turn out that our imbedding is essentially equivalent to the Skorohod imbedding.

Our task requires a more careful analysis of the waiting times $T_m(k)$ first. Recall the refinement property (41) of $B_m(t)$. Continuing that, we get

$$\begin{aligned} B_m(k2^{-2m}) &= B_{m+1}(2^{-2(m+1)}T_{m+1}(k)) = B_{m+2}(2^{-2(m+2)}T_{m+2}(T_{m+1}(k))) = \cdots \\ &= B_n(2^{-2n}T_n(T_{n-1}(\cdots(T_{m+1}(k))\cdots))), \end{aligned} \tag{66}$$

where $k \geq 0$ and $n > m \geq 0$. In other words, $B_n(t)$, $n > m$, visits the same dyadic points $k2^{-m}$ in the same order as $B_m(t)$, only the corresponding time instants can differ.

To simplify the notation, let us introduce

$$T_{m,n}(k) = T_n(T_{n-1}(\cdots(T_{m+1}(k))\cdots)) \quad (n > m \geq 0,\ k \geq 0).$$

Then (66) becomes

$$B_m(k2^{-2m}) = B_n(2^{-2n}T_{m,n}(k)) \quad (n > m \geq 0,\ k \geq 0). \tag{67}$$

The next lemma considers *time lags* of the form $2^{-2n}T_{m,n}(k) - k2^{-2m}$.

Note that in the proofs of the next two lemmas we make use of the following simple inequality, valid for arbitrary random variables Z_j, real numbers t_j, and events $A_j = \{Z_j \geq t_j\}$:

$$\begin{aligned} &\mathbf{P}\{Z_j \geq t_j \text{ for some } j \geq 1\} = \mathbf{P}\left\{\bigcup_{j=1}^{\infty} A_j\right\} \\ = \ &\mathbf{P}\{A_1\} + \mathbf{P}\{A_1^c A_2\} + \cdots + \mathbf{P}\left\{A_1^c \cdots A_j^c A_{j+1}\right\} + \cdots \\ \leq \ &\mathbf{P}\{Z_1 \geq t_1\} + \sum_{j=1}^{\infty} \mathbf{P}\{Z_j < t_j, Z_{j+1} \geq t_{j+1}\}. \end{aligned} \tag{68}$$

Lemma 5 *(a) For any $C \geq 3/2$, $K > 0$, and $m \geq 0$ take the following subset of the sample space:*

$$A_m = \left\{\max_{0 \leq k2^{-2m} \leq K} |2^{-2n}T_{m,n}(k) - k2^{-2m}| < \sqrt{18CKm}\, 2^{-m} \quad \text{for all } n > m\right\}. \tag{69}$$

Then for any $m \geq m_0(C, K)$ we have

$$\mathbf{P}\{A_m\} \geq 1 - 4(K2^{2m})^{1-C}. \tag{70}$$

(b) For any $K > 0$,

$$\max_{0 \leq k2^{-2m} \leq K} |2^{-2n} T_{m,n}(k) - k2^{-2m}| < \sqrt{27Km}\, 2^{-m}$$

with probability 1 for all $n > m$, for all but finitely many m.

Proof. Take any β, $1/2 < \beta < 1$, and $K' > K$; for example $\beta = 2/3$ and $K' = (4/3)K$ will do. Set $\alpha_j = 1 + \beta + \beta^2 + \cdots + \beta^j$ for $j \geq 0$,

$$Z_n = \max_{0 \leq k2^{-2m} \leq K} |T_{m,n}(k) - k2^{2(n-m)}|, \quad t_n = \alpha_{n-m-1}\sqrt{24CK'm}\, 2^{2n-m-2},$$

and $Y_{n+1} = \max_{0 \leq k2^{-2m} \leq K} |T_{n+1}(T_{m,n}(k)) - 4T_{m,n}(k)|$ for $n > m \geq 0$.

First, by Lemma 3(a),

$$\mathbf{P}\{Z_{m+1} \geq t_{m+1}\} = \mathbf{P}\left\{\max_{0 \leq k2^{-2m} \leq K} |T_{m+1}(k) - 4k| \geq \sqrt{24CK'm}\, 2^m\right\} \leq 2(K2^{2m})^{1-C}$$

if m is large enough.

Second, by the triangle inequality, $Z_{n+1} \leq 4Z_n + Y_{n+1}$ for any $n > m$. So

$$\mathbf{P}\{Z_n < t_n, Z_{n+1} \geq t_{n+1}\} \leq \mathbf{P}\{Z_n < t_n, Y_{n+1} \geq t_{n+1} - 4t_n\}.$$

If $Z_n < t_n$, then setting $j = T_{m,n}(k)$,

$$\begin{aligned} j2^{-2n} &< 2^{-2n}(k2^{2(n-m)} + t_n) = k2^{-2m} + \alpha_{n-m-1}\sqrt{24CK'm}\, 2^{-m-2} \\ &\leq K + 3\sqrt{2CKm}\, 2^{-m} \leq (4/3)K = K' \end{aligned}$$

holds for $m \geq m_0(C, K)$, since $\alpha_r < 1/(1-\beta) = 3$ (if $\beta = 2/3$) for any $r \geq 0$. Applying these first and Lemma 3(a) last, for $n > m \geq m_0(C, K)$ we get that

$$\begin{aligned} &\mathbf{P}\{Z_n < t_n, Z_{n+1} \geq t_{n+1}\} \\ &\quad\leq \mathbf{P}\left\{\max_{0 \leq j2^{-2n} \leq K'} |T_{n+1}(j) - 4j| \geq \sqrt{24CK'm}\, 2^n 2^{n-m}(\alpha_{n-m} - \alpha_{n-m-1})\right\} \\ &\quad\leq \mathbf{P}\left\{\max_{0 \leq j \leq K'2^{2n}} |T_{n+1}(j) - 4j| \geq \sqrt{24CK'n}\, 2^n\right\} \leq 2(K'2^{2n})^{1-C}. \end{aligned}$$

In the second inequality above we used that $\sqrt{m}2^{n-m}(\alpha_{n-m} - \alpha_{n-m-1}) = (2\beta)^{n-m}\sqrt{m} \geq \sqrt{n}$, which follows from the inequality $(2\beta)^{n-m} = (4/3)^{n-m} \geq \sqrt{1 + (n-m)/m}$, valid for any $n > m \geq 2$ (if $\beta = 2/3$).

Combining the results above,

$$
\begin{aligned}
&\mathbf{P}\left\{\max_{0\le k2^{-2m}\le K}|2^{-2n}T_{m,n}(k)-k2^{-2m}|\ge\sqrt{18CKm}\,2^{-m}\quad\text{for some } n>m\right\}\\
&= \mathbf{P}\left\{\max_{0\le k2^{-2m}\le K}|T_{m,n}(k)-k2^{2(n-m)}|\ge 3\sqrt{24CK'm}\,2^{2n-m-2}\quad\text{for some } n>m\right\}\\
&\le \mathbf{P}\left\{Z_n\ge t_n\quad\text{for some } n\ge m+1\right\}\\
&\le \mathbf{P}\left\{Z_{m+1}\ge t_{m+1}\right\}+\sum_{n=m+1}^{\infty}\mathbf{P}\left\{Z_n<t_n, Z_{n+1}\ge t_{n+1}\right\}\\
&\le \sum_{n=m}^{\infty}2(K2^{2n})^{1-C}=2(K2^{2m})^{1-C}\frac{1}{1-2^{2(1-C)}}\le 4(K2^{2m})^{1-C}
\end{aligned}
$$

if $C\ge 3/2$, say. This proves (a).

The statement in (b) follows by the Borel–Cantelli lemma. □

As (67) shows, $B_n(2^{-2n}T_{m,n}(k))=B_m(k2^{-2m})$ for any $k\ge 0$ and for any $n>m\ge 0$. A natural question, important particularly when looking at increments during short time intervals, is that how much time it takes for $B_n(t)$ to go from the point $B_m((k-1)2^{-2m})$ to its next "m-neighbor' $B_m(k2^{-2m})$. Is this time significantly different from 2^{-2m} for large values of m? Introducing the notation

$$
\tau_{m,n}(k)=T_{m,n}(k)-T_{m,n}(k-1)\quad(k\ge 1,\ n>m\ge 0),\tag{71}
$$

the nth time differences of the m-neighbors are $2^{-2n}\tau_{m,n}(k)$ $(k\ge 1)$. Note that $T_{m,n}(k)=\sum_{j=1}^{k}\tau_{m,n}(j)$, where, as can be seen from the construction and the argument below, the terms are independent and have the same distribution.

Let us look at $\tau_{m,n}(k)$ more closely. If $n=m+1$,

$$
\tau_{m,m+1}(k)=T_{m+1}(k)-T_{m+1}(k-1)=\tau_{m+1}(k),\tag{72}
$$

which is the double of a geometric random variable with parameter $p=1/2$, see (14). That is, $2^{-2(m+1)}\tau_{m+1}(k)$ is the length of the time period that corresponds to the time interval $[(k-1)2^{-2m},\ k2^{-2m}]$ after the next refinement of the construction.

Similarly, each unit in $\tau_{m+1}(k)$ will bring some $\tau_{m+2}(r)$ "offsprings" after the following refinement, and so on. Hence if $n>m$ is arbitrary, then given $T_{m,n}(k-1)=j$ for some integer $j\ge 0$, we have

$$
\tau_{m,n+1}(k)=T_{n+1}(j+\tau_{m,n}(k))-T_{n+1}(j)=\sum_{r=1}^{\tau_{m,n}(k)}\tau_{n+1}(j+r).\tag{73}
$$

For given $\tau_{m,n}(k)=s$ ($s>0$, even) its conditional distribution is the same as the distribution of a random variable T_s which is the double of a negative binomial random variable with parameters s and $p=1/2$, described by (19) and (20). Note that this conditional distribution of $\tau_{m,n+1}(k)$ is independent of the value of $T_{m,n}(k-1)$.

Though we will not explicitly use them, it is instructive to determine some further properties of a "prototype" $\tau_{m,n} = \tau_{m,n}(1)$. A recursive formula can be given for its expectation by (73) and the full expectation formula:

$$\mu_{n+1} = \mathbf{E}(\tau_{m,n+1}) = \mathbf{E}(\mathbf{E}(\tau_{m,n+1} \mid \tau_{m,n})) = \mathbf{E}(\tau_{m,n}\mathbf{E}(\tau_{n+1}(r))) = 4\mu_n.$$

Since $\mu_{m+1} = \mathbf{E}(\tau_{m+1}(r)) = 4$, it follows that

$$\mu_n = \mathbf{E}(\tau_{m,n}) = 2^{2(n-m)}.$$

This argument also implies that

$$\mathbf{E}(2^{-2(n+1)}\tau_{m,n+1} \mid 2^{-2n}\tau_{m,n}) = 2^{-2n}\tau_{m,n}.$$

These show that the sequence $(2^{-2n}\tau_{m,n})_{n=m+1}^{\infty}$ is a so-called *martingale*. Therefore a famous martingale convergence theorem [1, Section 5.4] implies that this sequence converges to a random variable t_m as $n \to \infty$, with probability 1, and t_m has finite expectation.

We mention that a similar recursion can be obtained for the variance that results

$$\mathbf{Var}(2^{-2n}\tau_{m,n}) < \frac{2}{3}2^{-4m}.$$

The next lemma gives an upper bound for the nth time differences of the m-neighbors by showing that during arbitrary many refinements, they cannot be "much" larger than $h = 2^{-2m}$, the original time difference of the m-neighbors. More accurately, they are less than a multiple of $h^{1-\delta}$, where $\delta > 0$ arbitrary.

Lemma 6 *(a) For any $K > 0$, δ such that $0 < \delta < 1$, and $C > 2/\delta$ we have*

$$\mathbf{P}\left\{\max_{1\le k2^{-2m}\le K} |2^{-2n}\tau_{m,n}(k) - 2^{-2m}| \ge 3C2^{-2m(1-\delta)} \quad \textit{for some } n > m\right\} \le \frac{K}{10}2^{-2m(\delta C-2)}.$$

(b) For any $K > 0$, and δ such that $0 < \delta < 1$,

$$\max_{1\le k2^{-2m}\le K} |2^{-2n}\tau_{m,n}(k) - 2^{-2m}| < \frac{7}{\delta}2^{-2m(1-\delta)},$$

with probability 1 for all $n > m$, for all but finitely many m.

Proof. This proof is very similar to the proof of Lemma 5. Take any β, $1/2 < \beta < 1$; for example $\beta = 2/3$ will do. Set $\alpha_j = 1 + \beta + \beta^2 + \cdots + \beta^j$ for $j \ge 0$. For any $m \ge 0$, consider an arbitrary k, $1 \le k \le K2^{2m}$. (The distribution of $\tau_{m,n}(k)$ does not depend on k.) Let

$$Z_n = |\tau_{m,n}(k) - 2^{2(n-m)}|, \quad t_n = \alpha_{n-m-1}C2^{2\delta m}2^{2(n-m)},$$

and $Y_{n+1} = |\tau_{m,n+1}(k)) - 4\tau_{m,n}(k)|$ for $n > m \ge 0$. We want to apply inequality (68).

First take $n = m+1$. By (72), $\frac{1}{2}\tau_{m,m+1}(k)$ is a geometric random variable with parameter $p = 1/2$. Then

$$\begin{aligned}
\mathbf{P}\left\{Z_{m+1} \geq t_{m+1}\right\} &= \mathbf{P}\left\{|\tau_{m,m+1}(k) - 4| \geq 4C2^{2\delta m}\right\} \\
&= \mathbf{P}\left\{\frac{1}{2}\tau_{m+1}(k) \geq 2 + 2C2^{2\delta m}\right\} < 2^{-4C\delta m},
\end{aligned}$$

because of the basic property of the tail of a geometric distribution.

Second, let $n > m$ be arbitrary. By the triangle inequality, $Z_{n+1} \leq 4Z_n + Y_{n+1}$. So we obtain

$$\begin{aligned}
&\mathbf{P}\left\{Z_n < t_n, Z_{n+1} \geq t_{n+1}\right\} \leq \mathbf{P}\left\{Z_n < t_n, Y_{n+1} \geq t_{n+1} - 4t_n\right\} \\
&\leq \sum_{s=1}^{2^{2(n-m)}+t_n} \mathbf{P}\left\{|T_s - 4s| \geq t_{n+1} - 4t_n \mid \tau_{m,n}(k) = s\right\} \ \mathbf{P}\left\{\tau_{m,n}(k) = s\right\} \\
&\leq \mathbf{P}\left\{\max_{1\leq s\leq 2^{2(n-m)}+t_n} |T_s - 4s|/\sqrt{8} \geq (t_{n+1} - 4t_n)/\sqrt{8}\right\} \ \sum_{s=1}^{\infty} \mathbf{P}\left\{\tau_{m,n}(k) = s\right\}, \qquad (74)
\end{aligned}$$

where we applied (73) and the conditional distribution of $\tau_{m,n+1}(k)$ mentioned there.

The sum in (74) is 1. Therefore we want to estimate the probability of the maximum there by using Lemma 2 with $N = 4C2^{2\delta m}2^{2(n-m)}$. This N is larger than $2^{2(n-m)} + t_n$ if m is large enough, depending on δ. (Remember that $\alpha_j < 3$ for any $j \geq 0$ if $\beta = 2/3$.) To apply Lemma 2 we have to compare $\sqrt{2CN\log N}$ to the right hand side of the inequality in (74):

$$\frac{\sqrt{2CN\log N}}{(t_{n+1} - 4t_n)/\sqrt{8}} = 2\left(\frac{(n-m)\log 4 + m\delta\log 4 + \log(4C)}{(4/3)^{2(n-m)}2^{2\delta m}}\right)^{1/2},$$

which is less than 1 for all $n > m$ if m is large enough, depending on δ.

Thus Lemma 2 gives that

$$\begin{aligned}
\mathbf{P}\left\{Z_n < t_n, Z_{n+1} \geq t_{n+1}\right\} &\leq \mathbf{P}\left\{\max_{1\leq s\leq N} |T_s - 4s|/\sqrt{8} \geq \sqrt{2CN\log N}\right\} \\
&\leq 2N^{1-C} = 2(4C2^{2\delta m}2^{2(n-m)})^{1-C}
\end{aligned}$$

for all $n > m$ as $m \geq m_0(\delta, C)$.

Summing up for $n > m$, we obtain the following estimate for any given k, $1 \leq k2^{-2m} \leq K$, as $m \geq m_0(\delta, C)$:

$$\begin{aligned}
&\mathbf{P}\left\{|2^{-2n}\tau_{m,n}(k) - 2^{-2m}| \geq 3C2^{-2m(1-\delta)} \quad \text{for some } n > m\right\} \\
&\leq \mathbf{P}\left\{Z_n \geq t_n \quad \text{for some } n \geq m+1\right\} \\
&\leq \mathbf{P}\left\{Z_{m+1} \geq t_{m+1}\right\} + \sum_{n=m+1}^{\infty} \mathbf{P}\left\{Z_n < t_n, Z_{n+1} \geq t_{n+1}\right\} \\
&\leq 2^{-4C\delta m} + 2^{2\delta(1-C)m}2(4C)^{1-C} \sum_{n=m+1}^{\infty} 2^{2(1-C)(n-m)} \\
&< (1/10)2^{-2m(\delta C-1)},
\end{aligned}$$

where we took into consideration that $0 < \delta < 1$, $C > 2$, $\delta C > 2$ by our assumptions.

Finally, the statement in (a) can be obtained by an application of the crude inequality (42),

$$\mathbf{P}\left\{\max_{1\le k\le K2^{2m}} |2^{-2n}\tau_{m,n}(k) - 2^{-2m}| \ge 3C2^{-2m(1-\delta)} \quad \text{for some } n > m\right\} \le K2^{2m}\frac{1}{10}2^{-2m(\delta C-1)},$$

as $m \ge m_0(\delta, C)$, which is equivalent to (a).

The statement in (b) follows by the Borel–Cantelli lemma with $C = 7/(3\delta)$ (say). □

Now we define a certain *imbedding* of shrunk random walks $B_m(k2^{-2m})$ into the Wiener process $W(t)$.

Lemma 7 *(a) For any $C \ge 3/2$, $K' > K > 0$, and any fixed $m \ge m_0(C, K, K')$ there exist random time instants $t_m(k) \in [0, K']$ such that*

$$\mathbf{P}\left\{W(t_m(k)) = B_m(k2^{-2m}), \quad 0 \le k2^{-2m} \le K\right\} \ge 1 - 4(K2^{2m})^{1-C},$$

where

$$\mathbf{P}\left\{\max_{0\le k2^{-2m}\le K} |t_m(k) - k2^{-2m}| \ge \sqrt{18CKm}\, 2^{-m}\right\} \le 4(K2^{2m})^{1-C}. \tag{75}$$

Moreover, if δ is such that $0 < \delta < 1$, $C > 2/\delta$, and $m \ge m_1(\delta, C, K, K')$, then we also have

$$\mathbf{P}\left\{\max_{1\le k2^{-2m}\le K} |t_m(k) - t_m(k-1) - 2^{-2m}| \ge 3C2^{-2m(1-\delta)}\right\} \le \frac{K}{10}2^{-2m(\delta C-2)} + 4(K2^{2m})^{1-C}. \tag{76}$$

(b) With probability 1, for any $K' > K > 0$, $0 < \delta < 1$, and for all but finitely many m there exist random time instants $t_m(k) \in [0, K']$ such that

$$W(t_m(k)) = B_m(k2^{-2m}) \quad (0 \le k2^{-2m} \le K),$$

where

$$\max_{0\le k2^{-2m}\le K} |t_m(k) - k2^{-2m}| \le \sqrt{27Km}\, 2^{-m},$$

and

$$\max_{1\le k2^{-2m}\le K} |t_m(k) - t_m(k-1) - 2^{-2m}| \le (7/\delta)2^{-2m(1-\delta)}.$$

Proof. By Lemma 5(a), fixing an $m \ge m_0(C, K, K')$, on a subset A_m of the sample space with $\mathbf{P}\{A_m\} \ge p_m = 1 - 4(K2^{2m})^{1-C}$, one has

$$\max_{0\le k2^{-2m}\le K} |2^{-2n}T_{m,n}(k) - k2^{-2m}| < \sqrt{18CKm}\, 2^{-m}, \tag{77}$$

for each $n > m$. In particular, the time instants $2^{-2n}T_{m,n}(k)$ are bounded from below by 0 and from above by $K + \sqrt{18CKm}\, 2^{-m} \le K'$. (Assume that $m_0(C, K, K')$ is chosen so.)

Applying a truncation $t^*_{m,n}(k) = \min\{K', 2^{-2n}T_{m,n}(k)\}$, for each k, $0 \le k2^{-2m} \le K$, we get a sequence in n bounded over the whole sample space, equal to the original one

for $\omega \in A_m$. It follows from the classical Weierstrass theorem [7, Section 2.42], that every bounded sequence of real numbers contains a convergent subsequence. To be definite, let us take the lower limit [7, Section 3.16] of the sequence:

$$t_m(k) = \liminf_{n\to\infty} t^*_{m,n}(k). \tag{78}$$

Then $t_m(k) \in [0, K']$.

By Theorem 3, with probability 1 the sample-functions of $B_n(t)$ uniformly converge to the corresponding sample-functions of the Wiener process that are uniformly continuous on [0, K']. (A continuous function on a closed interval is uniformly continuous [7, Section 4.19].) Thus (67) implies that for each k, $0 \le k2^{-2m} \le K$, we have $W(t_m(k)) = B_m(k2^{-2m})$, with probability at least p_m (on the set A_m where the truncated sequences coincide with the original ones).

To show it in detail, take any $\epsilon > 0$, any k ($0 \le k2^{-2m} \le K$), and a subsequence $t^*_{m,n_i}(k)$ converging to $t_m(k)$ as $i \to \infty$. Then

$$\begin{aligned}
& |W(t_m(k)) - B_m(k2^{-2m})| = |W(t_m(k)) - B_{n_i}(2^{-2n_i}T_{m,n_i}(k))| \\
\le\ & |W(t_m(k)) - W(2^{-2n_i}T_{m,n_i}(k))| + |W(2^{-2n_i}T_{m,n_i}(k)) - B_{n_i}(2^{-2n_i}T_{m,n_i}(k))| \\
<\ & \epsilon/2 + \epsilon/2 = \epsilon,
\end{aligned}$$

where the last inequality holds on the set A_m, for all but finitely many n_i. Since ϵ was arbitrary, it follows that $|W(t_m(k)) - B_m(k2^{-2m})| = 0$ on A_m.

Further, taking a limit in (77) with $n = n_i$ as $i \to \infty$ (on the set A_m), one obtains (75). Also, taking a similar limit in Lemma 6(a), $2^{-2n}\tau_{m,n_i}(k) \to t_m(k) - t_m(k-1)$ on the set A_m, and (76) follows.

The statements in (b) can be obtained similarly as in (a), applying Lemmas 5(b) and 6(b), or from (a) by the Borel–Cantelli lemma. □

We mention that for any $k \ge 0$ and $m \ge 0$, the sequence $2^{-2n}T_{m,n}(k)$ in fact converges to $t_m(k)$ with probability 1 as $n \to \infty$. However, a "natural" proof of this fact requires the martingale convergence theorem mentioned above, before Lemma 6, a tool of more advanced nature than the ones we use in this paper.

Next we want to show that the random time instants $s_m(k)$ of the Skorohod imbedding (65) and the $t_m(k)$'s defined in (78) are essentially the same. This requires a recollection of some properties of random walks.

We want to estimate the probability that with given positive integers j, x, u and r a random walk S_i goes from a point $|S_j| = x$ to $|S_{j+k}| = x + y$ so that $|S_{j+i}| < x + y$ while $1 \le i < k$ for some $y \le r$ and $k \ge u$, where k, y, and i are also positive integers.

The first passage distribution given in [2, Section III,7] can be applied here:

$$\mathbf{P}\{S_0 = 0, S_i < y\ (1 \le i < k), S_k = y\} = \frac{y}{k}\binom{k}{(k+y)/2}2^{-k}.$$

Hence, by Theorem 1,

$$\mathbf{P}\{|S_j| = x, |S_{j+i}| < x + y\ (1 \le i < k), |S_{j+k}| = x + y \ \text{ for some } \ y \le r\}$$

$$
\begin{aligned}
&\le \sum_{y=1}^{r} \frac{y}{k} \binom{k}{(k+y)/2} 2^{-k} \\
&\le (1+\epsilon)\frac{r}{k}\left(\Phi(r/\sqrt{k}) - \Phi(0)\right) \\
&\le (1+\epsilon)\frac{r}{k}\frac{r}{\sqrt{k}}\frac{1}{\sqrt{2\pi}} \le \frac{r^2}{k^{3/2}},
\end{aligned}
$$

where $\epsilon > 0$ is arbitrary, say equals 1, and $k \ge k_0$.

So the larger the value of k is, the smaller estimate of the probability we get. Thus for all positive integers j, x, r, and $u \ge k_0$,

$$
\mathbf{P}\left\{|S_j| = x, |S_{j+i}| < x + y \ (1 \le i < k), |S_{j+k}| = x + y \ \text{ for some } \ y \le r, \ k \ge u\right\} \le r^2/u^{3/2}, \tag{79}
$$

independently of the values of j and x.

Theorem 4 *The stopping times $s_m(k)$ ($k \ge 0$) of the Skorohod imbedding are equal to the time instants $t_m(k)$ of the imbedding defined in Lemma 7 on the set A_m of the sample space given by (69), with the possible exception of a zero probability subset.*

Therefore all statements in Lemma 7 hold when $s_m(k)$ replaces $t_m(k)$.

Proof. Fix an $m \ge m_0(C, K, K')$, where $m_0(C, K, K')$ is the same as in Lemma 7. Let the subset A_m of the sample space be given by (69).

Take $k = 1$ first. Since $s_m(1)$ is the smallest time instant where $|W(t)|$ is equal to 2^{-m}, and $|W(t_m(1))| = 2^{-m}$ on the set A_m, it follows that $s_m(1) \le t_m(1)$ on A_m. We want to show that on A_m the event $\{s_m(1) < t_m(1)\}$ has zero probability.

Indirectly, let us suppose that $\delta_m = t_m(1) - s_m(1) > 0$ on a subset C_m of A_m with positive probability. By (67), the first time instant where $|B_n(t)|$ equals $|B_m(2^{-2m})| = 2^{-m}$ is $2^{-2n}T_{m,n}(1)$ ($n > m$). So $|B_n(t)| < 2^{-m}$ if $0 \le t < 2^{-2n}T_{m,n}(1)$. On the other hand, by (55), $2^{-m} - n2^{-n/2} \le |B_n(s_m(1))| < 2^{-m}$ for $n \ge N_1(\omega)$ on a probability 1 ω-set. (Remember that $|W(s_m(1))| = 2^{-m}$.)

Since $\delta_m > 0$ on the set C_m, there exists an $N_2(\omega)$ such that $n2^{-n/2} < \delta_m/2$ for $n \ge N_2(\omega)$.

By (78), $t_m(1) = \liminf_{n\to\infty} 2^{-2n}T_{m,n}(1)$ on the set A_m. The properties of the lower limit [7, Section 3.17] imply that on the subset C_m there exists an $N_3(\omega)$ such that $2^{-2n}T_{m,n}(1) > t_m(1) - \delta_m/2$ for $n \ge N_3(\omega)$.

Set $N(\omega) = \max\{N_1(\omega), N_2(\omega), N_3(\omega)\}$ for $\omega \in C_m$. Since $B_n(t) = 2^{-n}\tilde{S}_n(t2^{2n})$, the statements above imply that on the set C_m the random walk $\tilde{S}_n(t)$ have the following properties for $n \ge N(\omega)$:

(a) $|\tilde{S}_n(s_m(1)2^{2n})| \ge 2^{n-m} - n2^{n/2}$,

(b) $|\tilde{S}_n(t)| < 2^{n-m}$ for $s_m(1)2^{2n} \le t < T_{m,n}(1)$, where $T_{m,n}(1) - s_m(1) > (\delta_m/2)2^{2n} > n2^{3n/2}$,

(c) $|\tilde{S}_n(T_{m,n}(1))| = 2^{n-m}$.

Let $D_{m,n}$ denote the subset of C_m on which (a), (b), and (c) hold for a fixed n. Since $D_{m,n} \subset D_{m,n+1}$ for each n, by the continuity property of probability [7, Section 11.3], we have $\lim_{n\to\infty} \mathbf{P}\{D_{m,n}\} = \mathbf{P}\{C_m\} > 0$. This implies that there exists an integer n_0 such that

$\mathbf{P}\{D_{m,n}\} \geq \frac{1}{2}\mathbf{P}\{C_m\} > 0$ holds for all $n \geq n_0$ (say). In other words, for all large enough values of n, the probability of the event that (a), (b), and (c) hold simultaneously is larger than a fixed positive number.

To get a contradiction, we apply (79) to $\tilde{S}_n(t)$, with $r = n2^{n/2}$ and $u = n2^{3n/2}$. Theorem 1, that was used to deduce (79), still applies since $r = o(u^{2/3})$, i.e. $r/\sqrt{u} = o(u^{1/6})$. Now the first passage time when $|\tilde{S}_n(t)|$ hits 2^{-2m} is $T_{m,n}(1)$. Thus the probability that $\tilde{S}_n(t)$ satisfies (a), (b), and (c) simultaneously is less than or equal to

$$\frac{r^2}{u^{3/2}} = \frac{(n2^{n/2})^2}{(n2^{3n/2})^{3/2}} = \frac{\sqrt{n}}{2^{5n/4}},$$

which goes to zero as $n \to \infty$. This contradicts the statement above that for all large enough value of n, the event that (a), (b), and (c) hold has a probability larger than a fixed positive number. This proves the lemma for $k = 1$: $s_m(1) = t_m(1)$ on the set A_m, with the possible exception of a zero probability subset.

For $k > 1$, one can proceed by induction. Assume that $s_m(k-1) = t_m(k-1)$ holds on A_m except possibly for a subset of probability zero. The proof that then $s_m(k) = t_m(k)$ holds as well is essentially the same as the proof of the case $k = 1$ above. It is true because on one hand $s_m(k)$ is defined recursively in (65), using $s_m(k-1)$, the same way as $s_m(1)$ is defined. On the other hand, by (71), $T_{m,n}(k) = T_{m,n}(k-1) + \tau_{m,n}(k)$, where the $\tau_{m,n}(k)$ is defined the same way as $\tau_{m,n}(1) = T_{m,n}(1)$. Also, remember that on the set A_m, $t_m(j) = \liminf_{n\to\infty} T_{m,n}(j)$ for $j = k-1$ or $j = k$. □

6 Some Properties of the Wiener Process

Theorem 3 above indicates that the sample-functions of the Wiener process are arbitrarily close to the sample-functions of $B_n(t)$ if n is large enough, with probability 1. The sample-functions of $B_n(t)$ are broken lines that have a chance of 1/2 to turn and have a corner at any multiple of time $1/2^{2n}$, so at more and more instants of time as $n \to \infty$. Moreover, the magnitude of the slopes of the line segments that make up the graph of $B_n(t)$ is

$$\frac{1/2^n}{1/2^{2n}} = 2^n \to \infty \quad \text{as} \quad n \to \infty.$$

Therefore one would suspect that the sample-functions of the Wiener process are typically nowhere differentiable. As we will see below, this is really true. Thus typical sample-functions of the Wiener process belong to the “strange” class of the everywhere continuous but nowhere differentiable functions.

Theorem 5 *With probability* 1*, the sample-functions of the Wiener process are nowhere differentiable.*

Proof. It suffices to show that with probability 1, the sample-functions are nowhere differentiable on any interval $[0, K]$. Put $K_0 = (3/2)K > 0$ (say). Then with probability

1, for all sample-functions and for all but finitely many m there exist time instants $t_m(k)$ $(0 \le k2^{-2m} \le K_0)$ with the properties described in Lemma 7(b). In particular,

$$\max_{0 \le k2^{-2m} \le K_0} t_m(k) \ge K_0 - \sqrt{27K_0 m}\, 2^{-m} > K$$

if m is large enough.

Fix an ω in this probability 1 subset of the sample space. This defines a specific sample-function of $W(t)$ and specific values of the random time instants $t_m(k)$. (To simplify the notation, in this proof we suppress the argument ω.) Then choosing an arbitrary point $t \in [0, K]$, for each m large enough, one has $t_m(k-1) \le t < t_m(k)$ for some k, $0 < k2^{-2m} \le K_0$. Taking for instance $\delta = 1/4$ in Lemma 7(b), we get $t_m(k) - t_m(k-1) \le 29 \cdot 2^{-(3/2)m}$ and

$$|W(t_m(k)) - W(t_m(k-1))| = |B_m(k2^{-2m}) - B_m((k-1)2^{-2m})| = 2^{-m}.$$

Set $t_m^* = t_m(k)$ if $|W(t) - W(t_m(k))| \ge |W(t) - W(t_m(k-1))|$ and $t_m^* = t_m(k-1)$ otherwise. Then $|W(t) - W(t_m^*)| \ge (1/2)2^{-m}$. So $|t_m^* - t| \le 29 \cdot 2^{-(3/2)m} \to 0$ and

$$\left| \frac{W(t_m^*) - W(t)}{t_m^* - t} \right| \ge \frac{(1/2)2^{-m}}{29 \cdot 2^{-(3/2)m}} = \frac{1}{58} 2^{m/2} \to \infty,$$

as $m \to \infty$. This shows that the given sample-function cannot be differentiable at any point $t \in [0, K]$. $\square$

It has important consequences in the definition of stochastic integrals that, as shown below, the graph of a typical sample-function of the Wiener process has infinite length. In general, (the graph of) a function f defined on an interval $[a, b]$ has *finite length* (or f is said to be of *bounded variation* on $[a, b]$) if there exists a finite constant c such that for any *partition* $a = x_0 < x_1 < \cdots < x_{n-1} < x_n = b$, the sum of the absolute values of the corresponding changes does not exceed c:

$$\sum_{j=1}^{n} |f(x_j) - f(x_{j-1})| \le c.$$

The smallest c with this property is called the *total variation* of f over $[a, b]$, denoted $V(f(t),\ a \le t \le b)$. Otherwise we say that the graph has *infinite length*, or f is of *unbounded variation* on $[a, b]$.

First let us calculate the total variation of a sample-function of $B_m(t)$ over an interval $[0, K]$. Each sample-function of $B_m(t)$ over $[0, K]$ is a broken line that consists of $K2^{2m}$ line segments with changes of magnitude 2^{-m}. So for any sample-function of $B_m(t)$,

$$V(B_m(t),\ 0 \le t \le K) = K2^{2m}2^{-m} = K2^m, \tag{80}$$

which tends to infinity as $m \to \infty$.

Lemma 8 *For any $K' > 0$, the sample-functions of the Wiener process over $[0, K']$ have infinite length (i.e. are of unbounded variation) with probability* 1.

Proof. By Lemma 7, for any $C \geq 3/2$, $K' > K > 0$, and $m \geq m_0(C, K, K')$ there exist time instants $t_m(k) \in [0, K']$ such that

$$\mathbf{P}\left\{W(t_m(k)) = B_m(k2^{-2m}), \quad 0 \leq k2^{-2m} \leq K\right\} \geq 1 - 4(K2^{2m})^{1-C}. \tag{81}$$

For each $m \geq 0$ define the following event:

$$C_m = \{V(W(t),\ 0 \leq t \leq K') < K2^m\}.$$

Then $C_m \subset C_{m+1}$ for any $m \geq 0$.

For any sample-function of $W(t)$, take the partition $0 = t_m(0) < t_m(1) < \cdots < t_m(K2^{2m})$. (To alleviate the notation, we suppress the dependence on ω.) By (81), for any $m \geq m_0(C, K, K')$, the sum of the corresponding absolute changes is equal to $K2^{2m}2^{-m} = K2^m$, with probability at least $1 - 4(K2^{2m})^{1-C}$.

This shows that then $\mathbf{P}\{C_m\} < 4(K2^{2m})^{1-C}$. Take the event

$$C_\infty = \{V(W(t),\ 0 \leq t \leq K') < \infty\}.$$

The continuity property of probability implies that $\mathbf{P}\{C_m\} \to \mathbf{P}\{C_\infty\}$ as $m \to \infty$, that is, $\mathbf{P}\{C_\infty\} = 0$. □

The next lemma shows a certain uniform continuity property of the Wiener process. An interesting consequence of the lemma is that for any $u > 0$ the probability that $|W(t) - W(s)| \geq u$ holds for some $s, t \in [0, K]$, $|t - s| \leq h$ can be made arbitrarily small if a small enough h is chosen. More accurately, the lemma shows that only with small probability can the increment of the Wiener process be larger than $c\sqrt{h}$ if the constant c is large enough. Now $\sqrt{h}$ is much larger than h for small values of h, so this also indicates why sample-functions of the Wiener process are not differentiable. At the same time it gives a rough measure of the so-called *modulus of continuity* of the process. Basically, the proof relies on Theorem 1a and Theorem 3.

Lemma 9 *For any $K > 0$, $0 < \delta < 1$, and $u > 0$ there exists an $h_0(K, \delta, u) > 0$ such that*

$$\mathbf{P}\left\{\max_{s,t\in[0,K],\ |t-s|\leq h} |W(t) - W(s)| \geq u\right\} \leq 7e^{-\frac{u^2}{2h}(1-\delta)}, \tag{82}$$

for all positive $h \leq h_0(K, \delta, u)$.

Proof. First we choose a large enough $C \geq 3/2$ such that $2/(C-1) < \delta/2$. For instance, $C = 1 + (6/\delta)$ will do.

By (54), the probability in (82) cannot exceed

$$6(K2^{2n})^{-6/\delta} + \mathbf{P}\left\{\max_{0\leq s\leq K-h} \max_{s\leq t\leq s+h} |B_n(t) - B_n(s)| \geq u - 2n2^{-n/2}\right\}, \tag{83}$$

for $n \geq n_0(K, \delta)$. (Remember that $1 - C = -6/\delta$ now.)

By definition, $B_n(t) = 2^{-n}\tilde{S}_n(t2^{2n})$ for $t \geq 0$. For each $s \leq t$ from $[0, K]$ and $n \geq n_0(K, \delta)$ take the integers $s_n = \lceil s2^{2n} \rceil$ and $t_n = \max\{s_n, \lfloor t2^{2n} \rfloor\}$. ($\lceil x \rceil$ denotes the smallest integer greater than or equal to x, while $\lfloor x \rfloor$ is the largest integer smaller than or equal to x.)

Then $|t_n - t2^{2n}| \leq 1$ and so $|\tilde{S}_n(t_n) - \tilde{S}_n(t2^{2n})| \leq 1$, similarly for s_n. Moreover, $0 \leq t_n - s_n \leq h2^{2n}$ if $0 \leq t - s \leq h$. Hence (83) does not exceed

$$6(K2^{2n})^{-6/\delta} + \mathbf{P}\left\{\max_{0 \leq j \leq K2^{2n}} \max_{0 \leq k \leq h2^{2n}} |\tilde{S}_n(j + k) - \tilde{S}_n(j)| \geq 2^n(u - 2n2^{-n/2}) - 2\right\}, \tag{84}$$

for $n \geq n_0(K, \delta)$.

The distribution of $\tilde{S}_n(j + k) - \tilde{S}_n(j)$ above is the same as the distribution of a random walk $S(k)$, for any value of $k \geq 0$, independently of $j \geq 0$. Also, the largest possible value of $|S(k)|$ is k. Therefore by Theorem 1a, the inequality (5), and the crude estimate (42),

$$\begin{aligned}
&\mathbf{P}\left\{\max_{0 \leq k \leq h2^{2n}} |\tilde{S}_n(j + k) - \tilde{S}_n(j)| \geq 2^n(u - 2n2^{-n/2} - 2 \cdot 2^{-n})\right\} \\
&\leq \mathbf{P}\left\{\max_{u\sqrt{1-\delta/2}\, 2^n \leq k \leq h2^{2n}} \frac{|S(k)|}{\sqrt{k}} \geq \frac{u}{\sqrt{h}}\sqrt{1 - \delta/2}\right\} \leq h2^{2n}e^{-\frac{u^2}{2h}(1-\delta/2)}.
\end{aligned}$$

Here it was assumed that $2n2^{-n/2} + 2 \cdot 2^{-n} \leq u(1 - \sqrt{1 - \delta/2})$, which certainly holds if $n \geq n_1(K, \delta, u) \geq n_0(K, \delta)$. Also, we assumed that $\frac{u}{\sqrt{h}}\sqrt{1 - \delta/2} \geq 3/\sqrt{2\pi}$, see (6), which is true if h is small enough, depending on δ and u.

Consequently, applying the crude estimate (42) again for (84), we obtain

$$\begin{aligned}
&\mathbf{P}\left\{\max_{s,t \in [0,K],\ |t-s| \leq h} |W(t) - W(s)| \geq u\right\} \\
&\leq 6(K2^{2n})^{-6/\delta} + K2^{2n}h2^{2n}e^{-\frac{u^2}{2h}(1-\delta/2)} \\
&= 6e^{-\frac{6}{\delta}(\log K + 2n\log 2)} + Khe^{4n\log 2 - \frac{u^2}{2h}(1-\delta/2)}.
\end{aligned}$$

Now we select an integer $n \geq n_1(K, \delta, u)$ such that $-\frac{6}{\delta}(\log K + 2n\log 2) \leq -\frac{u^2}{2h}$. The choice

$$n = \left\lceil \frac{1}{2\log 2}\left(\frac{u^2}{2h}\frac{\delta}{6} - \log K\right)\right\rceil$$

will do if h is small enough, $0 < h \leq h_0(K, \delta, u)$, so that $n \geq n_1(K, \delta, u) \geq 2$. Then $n \leq \frac{3}{2}\frac{1}{2\log 2}\left(\frac{u^2}{2h}\frac{\delta}{6} - \log K\right)$ holds as well.

With this n we have $4n\log 2 \leq \frac{u^2}{2h}\delta/2 + \log(K^{-3})$, and so

$$\begin{aligned}
&\mathbf{P}\left\{\max_{s,t \in [0,K],\ |t-s| \leq h} |W(t) - W(s)| \geq u\right\} \\
&\leq 6e^{-\frac{u^2}{2h}} + KhK^{-3}e^{-\frac{u^2}{2h}(1-\delta/2-\delta/2)} \\
&\leq (6 + h/K^2)e^{-\frac{u^2}{2h}(1-\delta)}.
\end{aligned}$$

If $K \geq 1$, then $h/K^2 \leq 1$ and (82) follows. If $K < 1$, the maximum in (82) cannot exceed the maximum over the interval $[0, 1]$. Then taking $h_0(K, \delta, u) = h_0(1, \delta, u)$, (82) follows again. $\Box$

7 A Preview of Stochastic Integrals

To show how stochastic integrals come as natural tools when working with differential equations including random effects, and what kind of problems arise when one wants to define them, let us start with the simplest ordinary differential equation

$$x'(t) = f(t) \quad (t \geq 0),$$

where f is a continuous function. If $x(0)$ is given, its unique solution can be obtained by integration,

$$x(t) - x(0) = \int_0^t f(s)\, ds \quad (t \geq 0).$$

Now we modify this simple model by introducing a random term, very customary in several applications:

$$x'(t) = f(t) + g(t)W'(t) \quad (t \geq 0),$$

where f and g are continuous random functions and $W'(t)$ is the so-called *white noise* process. Now we know from Theorem 5 that $W'(t)$ does not exist (at least not in the ordinary sense), but after integration we may get some meaningful solution,

$$x(t) - x(0) = \int_0^t f(s)\, ds + \int_0^t g(s)\, dW(s) \quad (t \geq 0).$$

The second integral here is what one wants to call a stochastic integral if it can be defined properly.

A natural idea to define such a stochastic integral is to define it as a *Riemann–Stieltjes integral* [7, Chapter 6] for each sample-function separately. It means that one takes partitions $0 = s_0 < s_1 < \cdots < s_{n-1} < s_n = t$, and Riemann–Stieltjes sums

$$\sum_{k=1}^{n} g(u_k)(W(s_k) - W(s_{k-1})),$$

where $u_k \in [s_{k-1}, s_k]$ is arbitrary. (We suppress the argument ω that would refer to a specific sample-function in order to alleviate the notation.) Then one would hope that as the norm of the partition $\|\mathcal{P}\| = \max_{1 \leq k \leq n} |s_k - s_{k-1}|$ tends to 0, the Riemann–Stieltjes sums converge to the same limit when fixing a specific point ω in the sample space.

One problem is that it cannot happen to all continuous random functions g. The reason is that $W(s)$ has unbounded variation over the interval $[0, t]$ —as we saw it in Lemma 8. The random function g could be chosen so that a Riemann–Stieltjes sum gets arbitrary close to the total variation, which is ∞. Naturally, this is the case with not only the Wiener process,

but with any process whose sample functions have unbounded variation, see e.g. [5, Section I.7].

But there is another problem connected to the choice of the points $u_k \in [s_{k-1}, s_k]$ in the Riemann–Stieltjes sums above. This choice unfortunately does matter, not like in the case of ordinary integration. The reason is again the unbounded variation of the sample-functions. The easiest way to illustrate it is using *discrete stochastic integrals*, that is, sums of random variables. (Such a sum is essentially the same as a Riemann–Stieltjes sum above.)

So let $S_0 = 0$, $S_n = \sum_{k=1}^{n} X_k$ is a (simple, symmetric) random walk, just like in Section 1. In the following examples S_n will play the role of the function $g(t)$ above, and the white noise process $W'(t)$ is substituted by the increments X_n. In the first case (that corresponds to an *Itô-type stochastic integral*), we define the discrete stochastic integral as $\sum_{k=1}^{n} S_{k-1}X_k$. Observe that in this case the integrand is always taken at the left endpoint of the subintervals. A usual reasoning behind this is that X_k gives the "new information" in each term, while the integrand S_{k-1} depends only on the past, that is, *non-anticipating*: independent of the future values $X_k, X_{k+1}, \ldots$.

This discrete stochastic integral can be evaluated explicitly as

$$\begin{aligned}\sum_{k=1}^{n} S_{k-1}X_k &= \sum_{k=1}^{n} S_{k-1}(S_k - S_{k-1}) \\ &= \frac{1}{2}\sum_{k=1}^{n}(S_k^2 - S_{k-1}^2) - \frac{1}{2}\sum_{k=1}^{n}(S_k - S_{k-1})^2 = \frac{S_n^2}{2} - \frac{n}{2}.\end{aligned}$$

Here we used that the first resulting sum telescopes and $S_0^2 = 0$, while each term $(S_k - S_{k-1})^2$ in the second resulting sum is equal to 1. The interesting feature of the result is that it contains the non-classical term $-n/2$. The "non-classical" phrase refers to the fact that $\int_0^{s_n} s\, ds = {s_n}^2/2$. Altogether, this formula is a special case of the important *Itô formula*, one of our main subjects from now on.

Of course, it is also interesting to see what happens if the integrand is always evaluated at the right endpoints of the subintervals:

$$\begin{aligned}\sum_{k=1}^{n} S_k X_k &= \sum_{k=1}^{n} S_k(S_k - S_{k-1}) \\ &= \frac{1}{2}\sum_{k=1}^{n}(S_k^2 - S_{k-1}^2) + \frac{1}{2}\sum_{k=1}^{n}(S_k - S_{k-1})^2 = \frac{S_n^2}{2} + \frac{n}{2}.\end{aligned}$$

Note that the non-classical term is $+n/2$ here.

Taking the arithmetical average of the two formulae above we obtain a *Stratonovich-type stochastic integral*, which does not contain a non-classical term:

$$\sum_{k=1}^{n} \frac{S_{k-1} + S_k}{2} X_k = \sum_{k=1}^{n} S(k - \frac{1}{2}) X_k = \frac{S_n^2}{2}.$$

On the other hand, this type of integral has other disadvantages compared to the Itô-type one, resulting from the fact that here the integrand is "anticipating", not independent of the future.

After showing these (and other) examples in a seminar, P. Révész asked the question if there is a general method to evaluate discrete stochastic integrals of the type $\sum_{k=1}^{n} f(S_{k-1})X_k$ in closed form, where f is a given function defined on the set of integers $\mathbf{Z}$. In other words, does there exist a discrete Itô formula in general? The answer is yes, and fortunately it is quite elementary to see.

But before turning to this, let us see the relationship of such a formula to an alternative way of defining certain stochastic integrals. This important type of stochastic integrals is $\int_0^K f(W(s))\, dW(s)$, where $K > 0$ and f is a continuously differentiable function. In other words, the integrand is a smooth function of the Wiener process. The traditional definition of the Itô-type integral in this case goes quite similarly to the Riemann–Stieltjes integral.

Take an arbitrary partition $\mathcal{P} = \{0 = s_0, s_1, \dots, s_{n-1}, s_n = K\}$ on the time axis, and a corresponding Riemann-Stieltjes sum, evaluating the function always at the left endpoints of the subintervals,

$$\sum_{k=1}^{n} f(W(s_{k-1}))\ (W(s_k) - W(s_{k-1})).$$

This sum is a random variable, corresponding to the given partition. It can be proved that these random variables converge e.g. *in probability* to a certain random variable I, as the norm of the partition goes to 0. This random variable I is then called the Itô integral. We mention that "in probability" convergence means that for any $\epsilon > 0$ there exists a $\delta > 0$ such that

$$\mathbf{P}\left\{ |I - \sum_{k=1}^{n} f(W(s_{k-1}))\ (W(s_k) - W(s_{k-1}))| \geq \epsilon \right\} < \epsilon,$$

as $\|\mathcal{P}\| < \delta$.

The alternative method that we will follow in this paper is better suited to the relationship between the Wiener process and random walks discussed above. Mathematically, it somewhat reminds a *Lebesgue–Stieltjes integral* [7, Chapter 11]. The idea is that we first take a dyadic partition on the spatial axis, each subinterval having the length 2^{-m}, where m is a non-negative integer. Then we determine the corresponding first passage times $s_m(1), s_m(2), \dots$ of the Skorohod imbedding as explained above. These time instants can be considered as a random partition on the time axis that in general depends on the considered sample-function.

By Lemma 7b and Theorem 4, with probability 1, for any $K' > 0$ and for all but finitely many m, each $s_m(k)$ lies in the interval $[0, K']$ and $W(s_m(k)) = B_m(k2^{-2m})$, $0 \leq k2^{-2m} \leq K$. The shrunk random walk $B_m(t)$ can be expressed in terms of ordinary random walks by (40) as $B_m(k2^{-2m}) = 2^{-m}\tilde{S}_m(k)$. Now our definition of the Itô integral will be

$$\lim_{m\to\infty} \sum_{k=1}^{K2^{2m}} f(W(s_m(k-1)))\ (W(s_m(k)) - W(s_m(k-1))). \tag{85}$$

We will show later that this sum, which can be evaluated for each sample-function separately, converges with probability 1. Our method will be to find an other form of this sum by a discrete Itô formula and to apply the limit to the equivalent form so obtained.

8 A Discrete Itô Formula

Let f be a function defined on the set of integers $\mathbf{Z}$. First we define *trapezoidal sums* of f by

$$T_{j=0}^{k}\ f(j) = \epsilon_k \left\{ \frac{1}{2} f(0) + \sum_{j=1}^{|k|-1} f(\epsilon_k j)\ + \frac{1}{2} f(k) \right\}, \tag{86}$$

where $k \in \mathbf{Z}$ (so k can be negative as well!) and

$$\epsilon_k = \begin{cases} \ \ 1 & \text{if } \ k > 0 \\ \ \ 0 & \text{if } \ k = 0 \\ -1 & \text{if } \ k < 0\ . \end{cases} \tag{87}$$

The reason behind the -1 factor when $k < 0$ is the analogy with integration: when the upper limit of the integration is less than the lower limit, one can exchange them upon multiplying the integral by -1.

The next statement that we will call a *discrete Itô formula* is a purely algebraic one. It is reflected by the fact that though we will apply it exclusively for random walks, the lemma holds for any numerical sequence $X_r = \pm 1$, irrespective of any probability assigned to them.

Lemma 10 *Take any function f defined on $\mathbf{Z}$, any sequence $X_r = \pm 1$ $(r \geq 1)$, and let $S_0 = 0$, $S_n = X_1 + X_2 + \cdots + X_n$ $(n \geq 1)$. Then the following statements hold:*

DISCRETE ITÔ FORMULA

$$T_{j=0}^{S_n}\ f(j) = \sum_{r=1}^{n} f(S_{r-1}) X_r + \frac{1}{2} \sum_{r=1}^{n} \frac{f(S_r) - f(S_{r-1})}{X_r},$$

and

DISCRETE STRATONOVICH FORMULA

$$T_{j=0}^{S_n}\ f(j) = \sum_{r=1}^{n} \frac{f(S_{r-1}) + f(S_r)}{2} X_r.$$

PROOF. By the definition of a trapezoidal sum,

$$T_{j=0}^{S_r}\ f(j) - T_{j=0}^{S_{r-1}}\ f(j) = X_r \frac{f(S_{r-1}) + f(S_r)}{2}, \tag{88}$$

since if $S_r - S_{r-1} = X_r$ equals 1, one has to add a term $(f(S_{r-1}) + f(S_r))/2$, while if $X_r = -1$, one has to subtract this term.

Since $X_r = \pm 1$, the right hand side of (88) can be written as

$$T_{j=0}^{S_r}\ f(j) - T_{j=0}^{S_{r-1}}\ f(j) = f(S_{r-1}) X_r + \frac{1}{2} \frac{f(S_r) - f(S_{r-1})}{X_r}. \tag{89}$$

By summing up (89), respectively (88), for $r = 1, 2, \ldots, n$ we obtain the statements of the lemma, since the sum telescopes and $T_{j=0}^{S_0} f(j) = 0$:

$$\sum_{r=1}^{n} \left(T_{j=0}^{S_r} f(j) - T_{j=0}^{S_{r-1}} f(j) \right) = T_{j=0}^{S_n} f(j).$$

□

We need a version of Lemma 10 that can be applied for shrunk random walks $B_m(t)$ as well. Therefore we define trapezoidal sums of a function f over an equidistant partition with points $x = j\Delta x$, where $\Delta x > 0$ and j changes over the set of integers $\mathbf{Z}$. Here the function f is assumed to be defined on the set of real numbers $\mathbf{R}$. So a corresponding trapezoidal sum is

$$T_{x=0}^{a} f(x) \Delta x = \epsilon_a \Delta x \left\{ \frac{1}{2} f(0) + \sum_{j=1}^{(|a|/\Delta x)-1} f(\epsilon_a j \Delta x) + \frac{1}{2} f(a) \right\}, \tag{90}$$

where a is assumed to be an integer multiple of Δx and ϵ_a is defined according to (87). In the sequel this definition will be applied with $\Delta x = 2^{-m}$. We write the corresponding version of Lemma 10 directly for shrunk random walks $B_m(t)$, though this lemma is of purely algebraic nature as well.

Lemma 11 *Take any function f defined on $\mathbf{R}$, any real $K > 0$, and fix a non-negative integer m. Consider shrunk random walks $B_m(r2^{-2m}) = 2^{-m}\tilde{S}_m(r)$ ($r \geq 0$). Then the following statements hold ($\Delta x = 2^{-m}$, $\Delta t = 2^{-2m}$):*

Itô case

$$\begin{aligned} T_{x=0}^{B_m(K_m)} f(x) \Delta x &= \sum_{r=1}^{\lfloor K/\Delta t \rfloor} f(B_m((r-1)\Delta t)) \left(B_m(r\Delta t) - B_m((r-1)\Delta t)\right) \\ &+ \frac{1}{2} \sum_{r=1}^{\lfloor K/\Delta t \rfloor} \frac{f(B_m(r\Delta t)) - f(B_m((r-1)\Delta t))}{B_m(r\Delta t) - B_m((r-1)\Delta t)} \Delta t, \end{aligned} \tag{91}$$

and

Stratonovich case

$$\begin{aligned} &T_{x=0}^{B_m(K_m)} f(x) \Delta x \\ &= \sum_{r=1}^{\lfloor K/\Delta t \rfloor} \frac{f(B_m((r-1)\Delta t)) + f(B_m(r\Delta t))}{2} (B_m(r\Delta t) - B_m((r-1)\Delta t)), \end{aligned} \tag{92}$$

where $K_m = \lfloor K/\Delta t \rfloor \Delta t$.

Proof. The proof is essentially the same as in case of Lemma 10, therefore omitted. □

Now recall Lemma 7b and Theorem 4. With probability 1, for any $K' > K$ and for all but finitely many m there exist random time instants $s_m(r) \in [0, K']$ (the first passage times of the Skorohod imbedding) such that $W(s_m(r)) = B_m(r\Delta t)$ and

$$\max_{0 \leq r\Delta t \leq K} |s_m(r) - r\Delta t| \leq \sqrt{27Km}\, 2^{-m}, \tag{93}$$

going to 0 as $m \to \infty$.

In this light the shrunk random walks $B_m(t)$ can be replaced by the Wiener process in (91) and (92). Then the first sum on the right hand side of (91) becomes exactly the one whose limit as $m \to \infty$ is going to be our definition of Itô integral by (85). Similarly, the right hand side of (92) is the one whose limit will be our definition of the Stratonovich integral.

The most important feature of Lemma 11 is that these limits can be evaluated in terms of limits of other, simpler sums. An other gain is that after performing the limits, we will immediately obtain the important Itô and Stratonovich formulae for the corresponding types of stochastic integrals.

9 Stochastic Integrals and the Itô formula

Theorem 6 *Let f be a continuously differentiable function on the set of real numbers* $\mathbf{R}$, *and $K > 0$. For $m \geq 0$ and $k \geq 0$ take the first passage times $s_m(k)$ of the Skorohod imbedding of shrunk random walks into the Wiener process as defined by (65). Then the sums below converge with probability* 1*:*

Itô integral

$$\int_0^K f(W(s))\, dW(s) = \lim_{m\to\infty} \sum_{r=1}^{K2^{2m}} f(W(s_m(r-1)))\, (W(s_m(r)) - W(s_m(r-1))), \tag{94}$$

and

Stratonovich integral

$$\begin{aligned}\int_0^K & f(W(s)) \circ dW(s) \\ &= \lim_{m\to\infty} \sum_{r=1}^{K2^{2m}} \frac{f(W(s_m(r-1))) + f(W(s_m(r)))}{2}\, (W(s_m(r)) - W(s_m(r-1))).\end{aligned} \tag{95}$$

For the corresponding stochastic integrals we have the following formulae as well:

Itô formula

$$\int_0^{W(K)} f(x)\, dx = \int_0^K f(W(s))\, dW(s) + \frac{1}{2}\int_0^K f'(W(s))\, ds, \tag{96}$$

and

Stratonovich formula

$$\int_0^{W(K)} f(x)\, dx = \int_0^K f(W(s)) \circ dW(s). \tag{97}$$

Proof. By the Itô case of Lemma 11 and the comments made after lemma, with probability 1, for all but finitely many m, we have the next equation for the sum in (94):

$$\begin{aligned}&\sum_{r=1}^{\lfloor K/\Delta t\rfloor} \frac{f(W(s_m(r-1))) + f(W(s_m(r)))}{2}\, (W(s_m(r)) - W(s_m(r-1))) \\ &= T_{x=0}^{W(s_m(\lfloor K/\Delta t\rfloor))} f(x)\, \Delta x - \frac{1}{2}\sum_{r=1}^{\lfloor K/\Delta t\rfloor} \frac{f(W(s_m(r))) - f(W(s_m(r-1)))}{W(s_m(r)) - W(s_m(r-1))}\Delta t,\end{aligned} \tag{98}$$

where $\Delta x = 2^{-m}$ and $\Delta t = 2^{-2m}$.

For $t \in [0, K]$ set $t_m = \lfloor t/\Delta t \rfloor \Delta t$. Then $|t - t_m| \leq \Delta t = 2^{-2m}$. By (93), $|t_m - s_m(\lfloor t/\Delta t \rfloor)| \leq \sqrt{27Km}\, 2^{-m}$ with probability 1 if m is large enough. This implies that

$$\max_{0 \leq t \leq K} |t - s_m(\lfloor t/\Delta t \rfloor)| \to 0 \tag{99}$$

with probability 1 as $m \to \infty$. Further, the sample functions of the Wiener process being uniformly continuous on $[0, K']$ with probability 1, one gets that then

$$\max_{0 \leq t \leq K} |W(t) - W(s_m(\lfloor t/\Delta t \rfloor))| \to 0 \tag{100}$$

as well.

Particularly, it follows that $W(s_m(\lfloor K/\Delta t \rfloor)) \to W(K)$ with probability 1 as $m \to \infty$. On the other hand, the trapezoidal sum $T_{x=0}^{a} f(x)\ \Delta x$ of a continuous function f is a Riemann sum corresponding to the partition $\left\{0, \frac{1}{2}\Delta x, \frac{3}{2}\Delta x, \ldots, a - \frac{3}{2}\Delta x, a - \frac{1}{2}\Delta x, a\right\}$. Therefore the trapezoidal sums converge to $\int_{x=0}^{a} f(x)\ dx$ as $\Delta x \to 0$. These show that for any $\epsilon > 0$,

$$\begin{aligned}
&\left| \int_0^{W(K)} f(x)\ dx - T_{x=0}^{W(s_m(\lfloor K/\Delta t \rfloor))} f(x)\ \Delta x \right| \\
\leq\ &\left| \int_0^{W(s_m(\lfloor K/\Delta t \rfloor))} f(x)\ dx - T_{x=0}^{W(s_m(\lfloor K/\Delta t \rfloor))} f(x)\ \Delta x \right| + \left| \int_{W(s_m(\lfloor K/\Delta t \rfloor))}^{W(K)} f(x)\ dx \right| \\
<\ &\epsilon/2 + \epsilon/2 = \epsilon
\end{aligned}$$

with probability 1 if m is large enough. That is, the trapezoidal sum in (98) tends to the corresponding integral with probability 1:

$$\lim_{m \to \infty} T_{x=0}^{W(s_m(\lfloor K/\Delta t \rfloor))} f(x)\ \Delta x = \int_0^{W(K)} f(x)\ dx. \tag{101}$$

Now let us turn to the second sum in (98). By the definition of the first passage times, $W(s_m(r)) - W(s_m(r-1)) = \pm 2^{-m} = \pm \Delta x$, which tends to 0 as $m \to \infty$. Hence

$$\frac{f(W(s_m(r))) - f(W(s_m(r-1)))}{W(s_m(r)) - W(s_m(r-1))} = \frac{f(W(s_m(r)) \mp \Delta x) - f(W(s_m(r)))}{\mp \Delta x}. \tag{102}$$

We want to show that this difference quotient gets arbitrarily close to $f'(W(r\Delta t))$ if m is large enough.

To this end, let us consider the following problem from calculus. If f is a continuously differentiable function, $x_m \to x$ and $\Delta x_m \to 0$ as $m \to \infty$, let us consider the difference of $f'(x)$ and $(f(x_m + \Delta x_m) - f(x_m))/\Delta x_m$. By the mean value theorem, the latter diference quotient is equal to $f'(u_m)$, where u_m lies between $x_m \to x$ and $x_m + \Delta x_m \to x$. Since f' is continuous, this implies that

$$\frac{f(x_m + \Delta x_m) - f(x_m))}{\Delta x_m} \to f'(x), \tag{103}$$

as $m \to \infty$.

In our present context, $x = W(t)$ and $x_m = W(s_m(\lfloor t/\Delta t \rfloor))$, where $0 \le t \le K$. Since the sample functions of $W(t)$ are continuous with probability 1, it follows from the max-min theorem that their ranges are contained in bounded intervals. Over such a bounded interval the function f' is uniformly continuous, therefore (99), (100), and (103) imply

$$\max_{0 \le t \le K} \left| f'(W(t)) - \frac{f(W(s_m(\lfloor t/\Delta t \rfloor)) \mp \Delta x) - f(W(s_m(\lfloor t/\Delta t \rfloor)))}{\mp \Delta x} \right| \to 0 \tag{104}$$

with probability 1 as $m \to \infty$. (Remember that now $\Delta x = 2^{-m}$ and $\Delta t = 2^{-2m}$.)

Particularly, for any $\epsilon > 0$, we have

$$\begin{aligned} & \max_{1 \le r\Delta t \le K} \left| \frac{f(W(s_m(r))) - f(W(s_m(r-1)))}{W(s_m(r)) - W(s_m(r-1))} - f'(W(r\Delta t)) \right| \\ = \quad & \max_{1 \le r\Delta t \le K} \left| \frac{f(W(s_m(r)) \mp \Delta x) - f(W(s_m(r)))}{\mp \Delta x} - f'(W(r\Delta t)) \right| < \frac{\epsilon}{3K} \end{aligned} \tag{105}$$

with probability 1 assuming m is large enough.

The function $f'(W(s))$ is continuous with probability 1, so its Riemann sums over $[0, K]$ converge to the corresponding integral as the norm of the partition tends to 0. Thus by (105),

$$\begin{aligned} & \left| \int_0^K f'(W(s))\, ds - \sum_{r=1}^{\lfloor K/\Delta t \rfloor} \frac{f(W(s_m(r))) - f(W(s_m(r-1)))}{W(s_m(r)) - W(s_m(r-1))} \Delta t \right| \\ \le \quad & \sum_{r=1}^{\lfloor K/\Delta t \rfloor} \left| f'(W(r\Delta t)) - \frac{f(W(s_m(r)) \mp \Delta x) - f(W(s_m(r)))}{\mp \Delta x} \right| \Delta t \\ + \quad & \left| \int_0^{K_m} f'(W(s))\, ds - \sum_{r=1}^{\lfloor K/\Delta t \rfloor} f'(W(r\Delta t)) \Delta t \right| + \left| \int_{K_m}^K f'(W(s))\, ds \right| \\ < \quad & \frac{\epsilon}{3K} K + \frac{\epsilon}{3} + \frac{\epsilon}{3} = \epsilon, \end{aligned}$$

with probability 1 if m is large enough. Here $K_m = \lfloor K/\Delta t \rfloor \Delta t$.

Therefore the second sum in (98) also tends to the corresponding integral with probability 1:

$$\lim_{m \to \infty} \frac{1}{2} \sum_{r=1}^{\lfloor K/\Delta t \rfloor} \frac{f(W(s_m(r))) - f(W(s_m(r-1)))}{W(s_m(r)) - W(s_m(r-1))} \Delta t = \frac{1}{2} \int_0^K f'(W(s))\, ds.$$

This proves that the defining sum of the Itô integral in (94) converges with probability 1 as $m \to \infty$, and for the limit we have Itô formula (96).

Also, by the Stratonovich case of Lemma 11 and the comments made after the lemma, with probability 1, for all but finitely many m, we have the following equation for the sum

in (95):

$$\sum_{r=1}^{\lfloor K/\Delta t\rfloor} \frac{f(W(s_m(r-1))) + f(W(s_m(r)))}{2}(W(s_m(r)) - W(s_m(r-1)))$$
$$= T_{x=0}^{W(s_m(\lfloor K/\Delta t\rfloor))} f(x)\Delta x.$$

We saw in (101) that this trapezoidal sum converges to the corresponding integral with probability 1 as $m \to \infty$. Therefore the defining sum of the Stratonovich integral in (95) converges as well, and for the limit we have formula (97). □

Since the Itô and Stratonovich formulae are valid for the usual definitions of the corresponding stochastic integrals as well, this shows that the usual definitions agree with the definitions given in this paper.

As we mentioned in a special case, the interesting feature of Itô formula (96) is that it contains the non-classical term $\frac{1}{2}\int_0^K f'(W(s))\,ds$. If g denotes an antiderivative of the function f, then the Itô formula can be written as

$$g(W(t)) - g(W(0)) = \int_0^t g'(W(s))\,dW(s) + \frac{1}{2}\int_0^t g''(W(s))\,ds,$$

or formally as the following non-classical chain rule for differentials:

$$dg(W(t)) = g'(W(t))\,dW(t) + \frac{1}{2}g''(W(t))\,dt.$$

We mention that other, more complicated versions of Itô formula can be proved by essentially the same method, see [8]. Also, as shown there, multiple stochastic integrals can be defined analogously as the stochastic integrals defined above.

Acknowledgements. I am obliged to Pál Révész from whom I learned many of the ideas underlying this paper. Also, I am grateful to Gábor Tusnády for his valuable help and to Tamás Linder for several useful comments. Special thanks are due to the Informatics Department of the University of Pisa and particularly to Vincenzo Manca for their hospitality while writing the last part of the paper.

References

[1] L. Breiman (1968). *Probability.* Addison-Wesley, Reading, Massachusetts.

[2] W. Feller (1968). *An Introduction to Probability Theory and Its Applications, Volume I: Third Edition.* J. Wiley, New York.

[3] W. Feller (1966). *An Introduction to Probability Theory and Its Applications, Volume II.* J. Wiley, New York.

[4] F.B. Knight (1981). *Essentials of Brownian Motion and Diffusion.* American Mathematical Society, Providence, R.I.

[5] P. Protter (1992). *Stochastic Integration and Differential Equations.* Springer, Berlin.

[6] P. Révész (1990). *Random Walk in Random and Non-Random Environments.* World Scientific, Singapore.

[7] W. Rudin (1977). *Principles of Mathematical Analysis: Third Edition.* McGraw-Hill, New York.

[8] T. Szabados (1990). A Discrete Itô's Formula. *Colloquia Mathematica Societatis János Bolyai* **57**. *Limit Theorems in Probability and Statistics, Pécs (Hungary) 1989*, 491-502. North-Holland, Amsterdam.